# DENSITY ESTIMATION IN BESOV SPACES VIA BLOCK THRESHOLDING

by

Eric Chicken
Purdue University

Technical Report # 01-12

Department of Statistics
Purdue University
West Lafayette, IN USA

September 2001

# Density Estimation in Besov Spaces via Block Thresholding

Eric Chicken
Purdue University

**Abstract**

For density estimation, block thresholding is very adaptive and efficient over a variety of general function spaces. By using block thresholding on kernel density estimators, the optimal minimax rates of convergence of the estimator to the true distribution are attained. This rate holds for large classes of densities residing in Besov spaces, including discontinuous functions with the number of discontinuities growing with sample size. The results hold for both convolution and wavelet kernel methods. Additionally, the proposed wavelet estimator is an improvement on previous estimators in that it simultaneously achieves both local and global optimal rates through careful choice of block length and a truncation parameter for the estimate's orthogonal series expansion.

## 1. Introduction

Wavelets have been shown to be very successful in density estimation. Specifically, they excel in the areas of spatial adaptivity, optimality, and low computational cost. Typically, this adaptivity is achieved through the use of term-by-term thresholding of wavelet coefficients, such as the VisuShrink method of Donoho and Johnstone (1994) for nonparametric estimation of a noisy signal. There, the noisy signal is transformed into empirical wavelet coefficients by the discrete wavelet transform, these coefficients are shrunk, or "denoised", by comparison with a specified thresholding rule, and the underlying function is estimated by applying the inverse discrete wavelet transform to these modified coefficients. This method is adaptive, i.e., it works well without knowing the exact amount of "smoothness" of the function ahead of time, and is within a logarithmic factor of the optimal minimax convergence rate over large classes of Besov functions. This optimal rate is measured in a global sense via the mean integrated squared error.

The earliest wavelet density estimators were linear in nature, introduced by Doukhan (1988) and Doukhan and Leon (1990). These linear estimators belong to the class of orthogonal series estimators first studied by Cencov (1962), who introduced the idea of relating the coefficients in the orthogonal series expansion of a density to the expected value of their corresponding basis functions. Kerkyacharian and Picard (1992) and Donoho et al. (1996) showed that these linear wavelet density estimators can achieve fast convergence rates when the density lies within Besov spaces.

*MSC2000*: Primary 62G07; secondary 62G20.
*Key words and phrases.* Wavelets, density estimation, a daptive, Besov class.

A drawback to linear wavelet density estimators is that they may be suboptimal, see Vidakovic (1995). Better density estimators can be found by introducing thresholded wavelet density estimators. This thresholding of the wavelet coefficients makes the new estimators nonlinear in nature. With these nonlinear density estimators, Donoho et al. (1996) attain a convergence rate that is within a logarithm of the optimal rate of $n^{-2s/(2s+1)}$. This rate is attained by the use of term-by-term thresholding.

Block thresholding has been shown to be superior to term-by-term thresholding in terms of convergence rates. For certain Sobolev spaces, Pensky (1999) has shown that block thresholding can result in the optimal rate of convergence without the logarithmic penalty term. Her blocks are large, each consisting of an entire resolution level of coefficients. For the more general Besov spaces, Hall et al. (1998) have set forth wavelet and convolution kernel estimators that also achieve the minimax optimal convergence rate without penalty through the use of block thresholding. Here, a specified number of coefficients within a resolution level is used as a block, rather than an entire resolution as in Pensky's case.

In both of these papers in the preceeding paragraph, the rate of convergence is in the global sense rather than a pointwise sense, i.e., it is measured via the mean integrated square error between the true function and its estimate. In this paper, an estimator similar to that of Hall et al. (1998) is proposed that not only achieves the same optimal rate over Besov spaces, but which attains the optimal local, or pointwise, convergence rate as well. This is achieved in main part through the careful choice of block size and a truncation parameter for the estimator's orthogonal series expansion.

Section 2 of this paper gives some background on wavelets and the function spaces of interest. The wavelet and convolution kernel density estimators and the theorems regarding their convergence rates are set forth in section 3, and their proofs are given in section 4.

## 2. Definitions and Notation

### *2.1 Wavelets*

The wavelets used in this paper are defined in terms of a multiresolution analysis. Starting with the space $L_2$ of real functions, decompose it into a series of nested spaces $V_l$, where

$$\ldots \supset V_{l+1} \supset V_l \supset V_{l-1} \supset \ldots,$$

$$\overline{\bigcup_l V_l} = L_2,$$

$$\bigcap_l V_l = \{0\},$$

and

$$f \in V_l \Leftrightarrow f(2\cdot) \in V_{l+1} \text{ for any } l.$$

Then, carefully choose a function $\phi$ such that for any integer $i$, the set of functions $\{\phi_{ij} | i, j \text{ integers}\}$ is an orthonormal basis for $V_i$,where

$$\phi_{ij} = 2^{i/2}\phi(2^i \cdot -j).$$

The function $\phi$ is called the scaling function or the "father" wavelet. Let $W_l$ be the orthogonal complement of $V_l$ in the space $V_{l+1}$. Then a space $V_l$ can be decomposed into subspaces $V_{l-1}$ and $W_{l-1}$:

$$\begin{array}{lllll} V_J & \to V_{J-1} & \to V_{J-2} & \to \quad \ldots & \to V_m \\ & \searrow W_{J-1} & \searrow W_{J-2} & \searrow \quad \ldots & \searrow W_m, \end{array}$$

or,

$$V_J = V_m \oplus \bigoplus_{i=m}^{J-1} W_i,$$

where $m < J$. In particular, the entire space $L_2$ can be written as

$$L_2 = V_m \oplus \bigoplus_{i=m}^{\infty} W_i$$

for any fixed $m$. It can be shown that each space $W_i$ is spanned by functions $\psi_{ij}$, where

$$\psi_{ij} = 2^{i/2}\psi(2^i \cdot -j),$$

and these "mother" wavelet functions can be constructed explicitly from the father wavelet $\phi$. Additionally, the father and mother wavelets are contructed so that

$$\int \phi = \int \phi^2 = \int \psi^2 = 1$$

and

$$\int \psi = 0.$$

Although it is assumed here that the collection of $\phi_{ij}$ are an orthonormal basis for $V_i$, this requirement can be loosened. It is only necessary that these functions form a Riesz basis.

Several types of wavelets have been constructed, but the best known are those of Daubechies (1992). In her construction of the functions $\phi$ and $\psi$, she uses $\phi$ that give rise to an orthonormal basis. Each collection of $\{\phi_{ij} | j \text{ an integer}\}$ and $\{\psi_{ij} | j \text{ an integer}\}$ is then an orthonormal basis for $V_i$ and $W_i$, respectively. By definition of the spaces $V_i$ and $W_i$, the functions $\phi_{ij}$ and $\psi_{ij}$ are orthogonal to each other as well. Additionally, Daubechies' method also results in compactly supported wavelets. Note that in Daubechies (1992), an alternate method of indexing the multiresolution analysis is used.

The wavelet functions created above can be used to represent functions in $L_2$. Let $f$ be any real function in $L_2$. The projection of $f$ onto the space $V_i$ is

$$\text{proj}_{V_i} f(x) = \sum_j \alpha_{ij}\phi_{ij}(x),$$

where

$$\alpha_{ij} = \langle f, \phi_{ij} \rangle = \int f\phi_{ij}$$

is the inner product of $f$ and $\phi_{ij}$. Likewise, the projection of $f$ onto the space $W_i$ is

$$\text{proj}_{W_i} f(x) = \sum_j \beta_{ij}\psi_{ij}(x),$$

where

$$\beta_{ij} = \langle f, \psi_{ij}\rangle = \int f\psi_{ij}$$

is the inner product of $f$ and $\psi_{ij}$. The function $f$ can then be written as

$$f(x) = \sum_j \alpha_{mj}\phi_{mj}(x) + \sum_{i=m}^{\infty}\sum_j \beta_{ij}\psi_{ij}(x)$$

for some fixed $m$.

## *2.2 Besov, Hölder, and other spaces*

We start by defining the Besov space $B^s_{p,q}$. For $0 < p, q \leq \infty$ and $s > 0$, a function $f$ is said to be in this space if its Besov norm is finite:

$$\|f\|_{B^s_{p,q}} < \infty,$$

where, for $0 < s \leq 1$,

$$\|f\|_{B^s_{p,q}} = \|f\|_{L^p} + \begin{cases} \left(\int_0^\infty \frac{1}{h}\left[\frac{1}{h^s}\|f(\cdot + h) - f(\cdot)\|_{L^p}\right]^q dh\right)^{\frac{1}{q}}, & q < \infty, \\ \sup\limits_{h>0} \frac{\|f(\cdot + h) - f(\cdot)\|_{L^p}}{h^s}, & q = \infty. \end{cases}$$

The usual $L_p$ norm is used here,

$$\|f\|_{L^p} = \left(\int |f|^p\right)^{\frac{1}{p}}.$$

For $s > 1$, $s = s^* + t, 0 < t \leq 1$, and $s^*$ the largest integer strictly less than $s$,

$$\|f\|_{B^s_{p,q}} = \sum_{m=0}^{s^*} \|f^{(m)}\|_{B^t_{p,q}}.$$

Roughly, a function $f$ in a Besov space $B^s_{p,q}$ has $s$ derivatives and is in $L_p$.

This paper considers functions whose Besov norms are bounded. For any $0 < M < \infty$, we define the Besov ball as:

$$B^s_{p,q}(M) = \left\{f : \|f\|_{B^s_{p,q}} \leq M\right\}.$$

As a measure of the local risk at a point, the local Hölder class $\Lambda^s(M, x_0, \delta)$ is used. For $0 < s \leq 1$,

$$\Lambda^s(M, x_0, \delta) = \{f : |f(x) - f(x_0)| \leq M|x - x_0|^s, x \in (x - \delta, x_0 + \delta)\}.$$

For $s > 1$

$$\Lambda^s(M, x_0, \delta) = \{f : |f^{(s^*)}(x) - f^{(s^*)}(x_0)| \leq M|x - x_0|^t, x \in (x - \delta, x_0 + \delta)\},$$

where $t = s - s^*$.

Also of interest with regards to Besov spaces are some inclusion properties. If $s > s'$ or $s = s'$ and $q \leq q'$, then

$$B^s_{pq} \subseteq B^{s'}_{pq'}. \tag{1}$$

If $p' > p$ and $s' = s - 1/p + 1/p'$, then

$$B^s_{pq} \subseteq B^{s'}_{p'q}. \tag{2}$$

See Härdle et al. (1996).

In section 3, the functions of interest lie in a subset of a Besov space. Additionally, it is further assumed that the functions are compactly supported and uniformly bounded. Following Hall et al. (1998) notation, define

$$F^s_{pq}(M, L) = \{f \in B^s_{pq} : \text{supp} f \in [-L, L], \|f\|_{B^s_{p,q}} \leq M\}. \tag{3}$$

The functions $f$ to be estimated can be written as a sum of two functions $f_1$ and $f_2$. The first function will be assumed to lie in the space $F^s_{2,\infty}(M, L)$. The second will be an irregular function that does not lie in the same space as $f_1$. This second function will lie in one of two spaces, denoted by Hall, Kerkyacharian and Picard as $P_{d,\tau,L}$ and $F^{s_1}_{(s+1/2)^{-1},\infty}(M, L)$.

$P_{d,\tau,L}$ is the set of piecewise polynomials of degree $d$, support in $[-L, L]$, and with the number of discontinuities no more than $\tau$. $F^{s_1}_{(s+1/2)^{-1},\infty}(M, L)$ is defined by (3) above.

Let $V_{d,\tau}(F^s_{2,\infty}(M, L))$ be all the functions $f$ that can be written as $f_1 + f_2$, where $f_1 \in F^s_{2,\infty}(M, L)$ and $f_2 \in P_{d,\tau,L}$. $\tilde{V}_{s_1}(F^s_{2,\infty}(M, L))$ will be the space of functions where $f_1$ is as above, and $f_2 \in F^{s_1}_{(s+1/2)^{-1},\infty}(M, L)$. The theorems presented in the next section will involve functions in these spaces intersected with $B_\infty(A)$, the set of all functions uniformly bounded by $A < \infty$.

Finally, wavelet coefficients for functions in Besov spaces have the property that the Besov norm of the function $f$ can be represented as a sequence norm in terms of its wavelet coefficients (see Meyer (1990)). If $f \in B^s_{p,q}$ and $\{\alpha_{mj}, \beta_{ij}\}$ is the collection of wavelet coefficents of $f$, then, when $p < \infty$

$$\|f\|_{B^s_{p,q}} = \left(\sum_j (\alpha_{mj})^p\right)^{\frac{1}{p}} + \begin{cases} \left(\sum_{i=0}^{\infty} \left[2^{i\left(s+\frac{1}{2}-\frac{1}{p}\right)} \left(\sum_j |\beta_{ij}|^p\right)^{\frac{1}{p}}\right]^q\right)^{\frac{1}{q}}, & q < \infty \\ \sup_{i \geq m} 2^{i\left(s+\frac{1}{2}-\frac{1}{p}\right)} \left(\sum_j | \beta_{ij} |^p\right)^{\frac{1}{p}}, & q = \infty. \end{cases} \tag{4}$$

When $p = \infty$, $(\sum |\cdot|^p)^{1/p}$ is replaced with the supremum over the summation index.

## 3. Density Estimation

### *3.1 The Kernel Functions*

Two types of kernel estimators will be examined in this section: wavelet kernels and convolution kernels. Following the notation of Hall et al. (1998), let $K(x, y)$ be a kernel function on $R^2$, and define

$$K_i(x, y) = 2^i K(2^i x, 2^i y), i = 0, 1, 2, \ldots .$$

Additionally, $K_i f$ will be the integral operator defined as

$$K_i f(x) = \int K_i(x, y) f(y) dy.$$

For independent, identically distributed random variables $X_1, X_2, \ldots, X_n$ from the distribution $f$, let

$$\hat{K}_i(x) = \frac{1}{n} \sum_{m=1}^{n} K_i(x, X_m).$$

Note that $\hat{K}_i(x)$ is an unbiased estimate of $K_i f(x)$ for all $x$:

$$\begin{aligned} E(\hat{K}_i(x)) &= E\left[\frac{1}{n} \sum_{m=1}^{n} K_i(x, X_m)\right] \\ &= E[K_i(x, X_1)] \\ &= \int K_i(x, y) f(y) dy \\ &= K_i f(x). \end{aligned}$$

In the convolution case, $K(x, y) = K(x - y)$. In the wavelet case,

$$K(x, y) = \sum_j \phi(x - j)\phi(y - j),$$

where $\phi$ is the father wavelet used in the context of a multiresolution analysis of Daubechies (1992).

Additionally, there will be several restrictions on the choice of $K$. First, there exists a $Q \in L^2$ (and hence in $L^1$) such that

$$|K(x, y)| \leq Q(x - y) \text{ for all } x \text{ and } y. \tag{5}$$

Next, $K$ must satisfy the moment condition of order $N$:

$$\int |x|^{N+1} Q(x) dx < \infty$$

$$\text{and} \tag{6}$$

$$\int K(x, y)(y - x)^k dy = \delta_{0k} \text{ for } k = 0, 1, \ldots, N.$$

Finally, $Q$ is compactly supported, say

$$Q(x) = 0 \text{ when } |x| > q_0. \tag{7}$$

Condition (5) implies (by Young's inequality) that

$$\|K_i f\|_p \leq \|Q\|_1 \|f\|_p \tag{8}$$

for all $p \geq 1$. Condition (6) is the usual assumption about the order of a kernel. Condition (7) is presented only to simplify the proof. The conditions (5), (6) and (7) are met in the wavelet case if the mother wavelet $\psi$ has $N$ vanishing moments

$$\int x^k \psi_{ij}(x)dx = 0, k = 0, 1, \dots, N,$$

and if $\phi$ and $\psi$ are both bounded. See Kerkyacharian and Picard (1992).

Hall et al. (1998) defined their "innovation" kernel as

$$D_i(x, y) = K_{i+1}(x, y) - K_i(x, y)$$

for $i = 0, 1, \dots$. Let $D_i f$ be the integral operator $K_{i+1} f - K_i f$. Then, similarly to $\hat{K}_i$, an unbiased estimator of $D_i f(x)$ is

$$\hat{D}_i(x) = \frac{1}{n} \sum_{m=1}^{n} D_i(x, X_m).$$

In the wavelet case, $K$ and $D_i$ can be associated with the projection operators of the multiresolution analysis. $K(x, y)$ is the projection operator on to the space spanned by $\phi$ and its integer translates. In the notation of multiresolution analysis, this is the "coarse" space $V_0$. $D_i(x, y)$ is, then, the operator projecting on to the "detail" spaces $W_i$ of multiresolution analysis. The number of projections on to these detail spaces to be used will be finite, say $R$.

$K$ and $D_i$ perform similar tasks in the convolution case: namely, projection operators on to coarse and detail spaces. This innovation kernel will be used to define the density estimator in the next section.

### *3.2 The Density Estimator*

The density to be estimated may be written as

$$f(x) = K_0 f(x) + \sum_{i=0}^{\infty} D_i f(x). \tag{9}$$

The linear part, $K_0 f(x)$, will be estimated by $\hat{K}_0(x)$. The remaining part will estimated using thresholding methods, and hence is nonlinear in nature. The index $i$ will be truncated to some finite value $R$.

To understand the thresholding method, the wavelet case will be examined first. Let $\phi$ and $\psi$ be bounded father and mother wavelets of the multiresolution analysis satisfying conditions (5), (6), and (7). Define the dilations and translations of these two functions as

$$\phi_j(x) = \phi_{0j}(x) = \phi(x - j)$$

and

$$\psi_{ij}(x) = 2^{\frac{i}{2}} \psi(2^i x - j).$$

Let $\alpha_j =< f, \phi_j >$ and $\beta_{ij} =< f, \psi_{ij} >$ be the usual inner products as defined in section 2. Unbiased estimates of $\alpha_j$ and $\beta_{ij}$ are

$$\hat{\alpha_j} = \frac{1}{n} \sum_{m=1}^{n} \phi_j(X_m)$$

and

$$\hat{\beta_{ij}} = \frac{1}{n} \sum_{m=1}^{n} \psi_{ij}(X_m).$$

The linear part $K_0 f(x)$ can be written as

$$\begin{aligned} K_0 f(x) &= \int K(x,y) f(y) dy \\ &= \int \sum_j \phi(x-j)\phi(y-j) f(y) dy \\ &= \sum_j \phi(x-j) \int \phi(y-j) f(y) dy \\ &= \sum_j \alpha_j \phi_j(x). \end{aligned}$$

The estimate of $K_0 f(x)$ is then

$$\begin{aligned} \hat{K_0}(x) &= \sum_j \hat{\alpha_j} \phi_j(x) \\ &= \sum_j \phi(x-j) \left( \frac{1}{n} \sum_{m=1}^{n} \phi(X_m - j) \right) \\ &= \frac{1}{n} \sum_{m=1}^{n} \sum_j \phi(x-j)\phi(X_m - j) \\ &= \frac{1}{n} \sum_{m=1}^{n} K(x, X_m). \end{aligned}$$

Similarly, note that $K_{i+1}(x,y) - K_i(x,y)$, $i \geq 0$, is the projection operator onto the detail space $W_i$. From Vidakovic (1995), the projection onto $W_i$ can also be written as

$$\sum_j \psi_{ij}(x)\psi_{ij}(y),$$

so

$$D_i(x,y) = K_{i+1}(x,y) - K_i(x,y) = \sum_j \psi_{ij}(x)\psi_{ij}(y).$$

Therefore, in a manner similar to that used above on $K_0 f(x)$,

$$D_i f(x) = \sum_j \beta_{ij} \psi_{ij}(x).$$

The estimate of the nonlinear part $D_i f(x)$ is, then,

$$\hat{D}_i(x) = \sum_j \hat{\beta}_{ij}\psi_{ij}(x).$$

In the wavelet case, we can then rewrite (9) as

$$f(x) = \sum_j \alpha_j \phi_j(x) + \sum_{i=0}^{\infty}\sum_j \beta_{ij}\psi_{ij}(x), \tag{10}$$

and estimate (10) as

$$\begin{aligned}\hat{f}(x) =& \sum_j \hat{\alpha}_j \phi_j(x) + \sum_{i=0}^{R}\sum_j \hat{\beta}_{ij}\psi_{ij}(x) \qquad (11)\\ =& \hat{K}_0(x) + \sum_{i=0}^{R} \hat{D}_i(x)\end{aligned}$$

where $R$ is a finite truncation value for the infinite series. Thresholding will now be applied to the nonlinear part $\sum_i \sum_j \hat{\beta}_{ij}\psi_{ij}(x)$. The variance of $\hat{\beta}_{ij}\psi_{ij}(x)$ is $n^{-1}\text{var}\ (\psi_{ij}(X))\psi_{ij}^2(x)$ and the squared bias when leaving out the term associated with $\hat{\beta}_{ij}$ in (11) is $\beta_{ij}^2\psi_{ij}^2(x)$. It seems reasonable to keep the term $\hat{\beta}_{ij}\psi_{ij}(x)$ whenever the squared bias for removing that term is greater than its variance. Thus, $\hat{\beta}_{ij}\psi_{ij}(x)$ would be replaced by $\hat{\beta}_{ij}\psi_{ij}(x)I(\beta_{ij}^2 > cn^{-1})$ for some constant $c$. But, since $\beta_{ij}$ is unknown, $\hat{\beta}_{ij}\psi_{ij}(x)I(\hat{\beta}_{ij}^2 > cn^{-1})$ will be used. This term-by-term thresholding method of (11) leads to the following estimate of $f$:

$$\hat{f}(x) = \sum_j \hat{\alpha}_j \phi_j(x) + \sum_{i=0}^{R}\sum_j \hat{\beta}_{ij}\psi_{ij}(x)I(\hat{\beta}_{ij}^2 > cn^{-1}). \tag{12}$$

However, this term-by-term thresholding estimate results in poor mean squared error. The optimal rate can never be achieved. In fact, the best rate that can be attained is $n^{-s'}$ for some $s'$ strictly less than $2s/(2s+1)$. By using $c\log(n)/n$ rather than $cn^{-1}$ as a threshold, the convergence rate of (12) can be improved to a constant multiple of $(n^{-1}\log(n))^{2s/(2s+1)}$, which is still less than the optimal minimax rate. See Hall et al. (1998).

Instead of using this term-by-term thresholding method and the estimate at (12), block thresholding will be used to create a new density estimtor. In each resolution level $i$, the indices $j$ are divided up into nonoverlapping blocks of length $l$. Within this block, the average estimated squared bias $l^{-1}\sum_{j\in B(k)}\hat{\beta}_{ij}^2$ will be compared to the threshold. Here, $B(k)$ refers to the set of indices $j$ in block $k$. By estimating all of these squared coefficients together, the additional information allows a better comparison to the threshold, and hence a better convergence rate. If the average squared bias is larger than the threshold, all coefficients in the block will be kept. Otherwise, all coefficients will discarded.

Letting

$$B_{ik} = l^{-1} \sum_{j \in B(k)} \beta_{ij}^2$$

and estimating this with

$$\hat{B}_{ik} = l^{-1} \sum_{j \in B(k)} \hat{\beta}_{ij}^2,$$

the wavelet-based estimate of $f$ to be used in this paper becomes

$$\hat{f}(x) = \sum_j \hat{\alpha}_j \phi_j(x) + \sum_{i=0}^{R} \sum_k \sum_{j \in B(k)} \hat{\beta}_{ij} \psi_{ij}(x) I(\hat{B}_{ik} > cn^{-1}). \tag{13}$$

The non-wavelet, convolution kernel case will be treated similarly. First, note that

$$\begin{aligned} B_{ik} &= l^{-1} \sum_{j \in B(k)} \beta_{ij}^2 \\ &= l^{-1} \sum_{j \in B(k)} \int_{x:\psi_{ij}(x) \neq 0} \beta_{ij}^2 \psi_{ij}^2(x) dx \\ &= l^{-1} \int_{J_{ik}} \left( \sum_{j \in B(k)} \beta_{ij} \psi_{ij}(x) \right)^2 dx \\ &= l^{-1} \int_{J_{ik}} \left( D_{ik} f(x) \right)^2 dx, \end{aligned}$$

where

$$D_{ik} f(x) = \sum_{j \in B(k)} \beta_{ij} \psi_{ij}(x),$$

and

$$J_{ik} = \bigcup_{j \in B(k)} \{x : \psi_{ij}(x) \neq 0\} = \bigcup_{j \in B(k)} \{\text{supp } \psi_{ij}\}.$$

By a like argument,

$$\begin{aligned} \hat{B}_{ik} &= l^{-1} \sum_{j \in B(k)} \hat{\beta}_{ij}^2 \\ &= l^{-1} \int_{J_{ik}} \hat{D}_{ik}^2(x) dx, \end{aligned}$$

where

$$\hat{D}_{ik}(x) = \sum_{j \in B(k)} \hat{\beta}_{ij} \psi_{ij}(x).$$

The size of the interval $J_{ik}$ is $D2^{-i}l$ for some constant $D$ which depends on the length of the support and the amount of overlap of $\psi_{ij}$ since, with the exception of the Haar wavelet, the support of wavelet functions overlap one another. The wavelet density estimator (13) may then be written as

$$\hat{f}(x) = \hat{K}_0(x) + \sum_{i=0}^{R}\sum_{k} \hat{D}_{ik}(x) I(x \in J_{ik}) I(\hat{B}_{ik} > cn^{-1}).$$

If the support of the $\psi_{ij}$ were nonoverlapping, then the length of $J_{ik}$ would be $D'2^{-i}l$, where $D'$ depends only on the length of the support of the $\psi_{ij}$. Furthermore,

$$B_{ik} = l^{-1}\int_{J_{ik}} (D_{ik}f(x))^2\, dx = l^{-1}\int_{J_{ik}} (D_i f(x))^2 dx, \tag{14}$$

and

$$\hat{B}_{ik} = l^{-1}\int_{J_{ik}} (D_{ik}f(x))^2\, dx = l^{-1}\int_{J_{ik}} \hat{D}_i^2(x) dx, \tag{15}$$

since the $J_{ik}$ would only include the supports of the $\psi_{ij}$ in the $k$th block. The nonoverlapping wavelet density estimator (13) may then be written as

$$\hat{f}(x) = \hat{K}_0(x) + \sum_{i=0}^{R}\sum_{k} \hat{D}_i(x) I(x \in J_{ik}) I(\hat{B}_{ik} > cn^{-1}).$$

This alternate form of (13) is the model for the convolution kernel estimator. Replace the intervals $J_{ik}$ with nonoverlapping intervals $I_{ik}$ of length $2^{-i}l$. Then analogously to (14) and (15), define

$$A_{ik} = l^{-1}\int_{I_{ik}} (D_i f(x))^2 dx,$$

and estimate $A_{ik}$ with

$$\hat{A}_{ik} = l^{-1}\int_{I_{ik}} \hat{D}_i^2 f(x) dx.$$

The convolution kernel block thresholded equivalent of (13) is then

$$\hat{f}(x) = \hat{K}_0(x) + \sum_{i=0}^{R}\sum_{k} \hat{D}_i(x) I(x \in I_{ik}) I(\hat{A}_{ik} > cn^{-1}). \tag{16}$$

### *3.3 Convergence Rates for the Density Estimators*

The optimal minimax rate of convergence of an estimate of a density in a Besov space to the true underlying density is $O(n^{-2s/(2s+1)})$ for a function with unknown smoothness parameter $s$. For the wavelet kernel density estimator (13) and the appropriate choice of block length $l$ and series truncation parameter $R$, this rate is achieved over the space $\tilde{V}_{s_1}(F^s_{2,\infty}(M, L)) \cap B_\infty(A)$ as defined in section 2.

**Theorem 1** *Let $\hat{f}$ be the wavelet kernel density estimator (13). Let the block length $l$ be $\log n$, $R = \lfloor \log_2 n^{1-\varepsilon} \rfloor$ for some fixed $\varepsilon \in (0, 1/2]$ and suppose $\phi$ and $\psi$ are bounded. If $s_1 - s > s/\left((2s+1)(1-\varepsilon)\right)$, $\psi$ has $N-1$ vanishing moments, and*

$$c = A(0.08)^{-1}\left(C_2\|Q\|_2 + \|Q\|_1\sqrt{\frac{2N}{C_1(2N+1)}}\right)^2,$$

*where $C_1$ and $C_2$ are the universal constants from Talagrand (1994), then there exists a positive constant $C$ such that for all $1/2 < s < N$,*

$$\sup_{f \in \tilde{V}_{s_1}(F^s_{2,\infty}(M,L)) \cap B_\infty(A)} E\|\hat{f} - f\|_2^2 \leq Cn^{-2s/(2s+1)}.$$

The convolution kernel estimator (16) also achieves the global, optimal minimax convergence rate with this smaller block length, although over a different space of irregular Besov functions.

**Theorem 2** *Let $\hat{f}$ be the convolution kernel density estimator (16). Let $\tau_n$ be a sequence of positive numbers such that for all $\zeta > 0$, $\tau_n = O(n^{\zeta + 1/(2N+1)})$. Let the block length $l$ be $\log n$ and $R = \lfloor \log_2 n^{1-\varepsilon} \rfloor$, where $\varepsilon = \rho[(2N+1)(2(N-\rho)+1)]^{-1}$, and $\rho$ is any fixed number such that $0 < \rho < N - 1/2$. Let*

$$c = A(0.08)^{-1}\left(C_2\|Q\|_2 + \|Q\|_1\sqrt{\frac{2(N-\rho))}{C_1(2(N-\rho)+1)}}\right)^2,$$

*where $C_1$ and $C_2$ are the universal constants from Talagrand (1994). If $K$ satisfies (5), (6) with order $N-1$, and (7), and $1/2 < s < N - \rho$, then there exists a positive constant $C$ such that*

$$\sup_{d<N, \tau \leq \tau_n} \sup_{f \in V_{d\tau}(F^s_{2,\infty}(M,L)) \cap B_\infty(A)} E\|\hat{f} - f\|_2^2 \leq Cn^{-2s/(2s+1)}.$$

These two theorems differ from Hall et al. (1998) in that the block length $l$ is $\log n$ instead of $(\log n)^2$ and that their truncation parameter $R$ does not contain the exponent $1-\varepsilon$. Additionally, in theorem 1, the range of the functions covered by their estimator is slightly larger, and in theorem 2, the range of the unknown smoothness parameter has been lessened.

In the wavelet kernel case, the only restriction on the irregular part $f_2$ of the function $f$ in Hall, Kerkyacharian and Picard's paper is that $s_1 - s > s/(2s+1)$. Although the range of $\varepsilon$ extends to $1/2$ in theorem 1, to increase the scope of the spaces under consideration to be closer in size to that of Hall et al. (1998) it is advantageous to choose $\varepsilon$ to be near zero.

In the convolution kernel case, theorem 2 has reduced the range of the smoothness paramater $s$ found in Hall, Kerkyacharian and Picard from $(1/2, N)$ to $(1/2, N-\rho)$ for an arbitrarily small constant $\rho > 0$. However, if the reader is unhappy with the upper limit of $N - \rho$, and insists on $N$, this can be overcome by choosing a kernel with $N$

vanishing moments instead of $N-1$. Or, one may choose $\rho$ to be very small. However, by changing the value of $\rho$, the constant $C = C(\rho)$ will increase.

Another way of dealing with reduction in the range of $s$ in theorem 2 is by modifying the number of discontinuities. In Hall, Kerkyacharian and Picard, $\tau_n$ is a sequence such that for all $\zeta > 0$, $\tau_n = O(n^{\zeta+1/(2N+1)})$. By changing $\tau_n$ to a sequence of order $O(n^{1/(2N+1)-\zeta})$ for some fixed $\zeta \in (0, 1/(2N+1))$, the range of $s$ is restored to $(1/2, N)$. In this case, theorem 2 becomes:

**Theorem 3** *Let $\hat{f}$ be the convolution kernel density estimator (16). Let $\tau_n$ be a sequence of positive numbers such that for a fixed $\zeta \in (0, 1/(2N+1))$, $\tau_n = O(n^{1/(2N+1)} - \zeta)$. Let the block length $l$ be $\log n$ and $R = \lfloor \log_2 n^{1-\zeta} \rfloor$. Let*

$$c = A(0.08)^{-1}\left(C_2\|Q\|_2 + \|Q\|_1\sqrt{\frac{2N}{C_1(2N+1)}}\right)^2,$$

*where $C_1$ and $C_2$ are the universal constants from Talagrand (1994). If $K$ satisfies (5), (6) with order $N-1$, and (7), and $1/2 < s < N$, then there exists a positive constant $C$ such that*

$$\sup_{d<N, \tau \le \tau_n} \quad \sup_{f \in V_{d\tau}(F^s_{2,\infty}(M,L)) \cap B_\infty(A)} E\|\hat{f} - f\|_2^2 \le Cn^{-2s/(2s+1)}.$$

The reader is then left to decide between theorems 2 and 3 as to which is more beneficial to the problem at hand: larger range of adaptivity for the unknown smoothness parameter $s$, or the ability of the estimator to handle a larger number of discontinuities.

In all of the preceeding theorems on densities, the rate of convergence of Hall, Kerkyacharian and Picard is not affected by changing the block length, and indeed, there is an advantage to using this smaller block length in regards to local adaptivity.

Theorems 1, 2 and 3, show that the estimators (13) and (16) are globally adaptive in terms of the smoothness parameter $s$. The local adaptivity of a function at a point $x_0$ is determined by

$$E(\hat{f}(x_0) - f(x_0))^2.$$

As a measure of the local risk at a point, the local Hölder class $\Lambda^s(M, x_0, \delta)$ defined in section 2 is used. To achieve local adaptivity, Brown and Low (1996) showed that there is a penalty sufferred. Namely, a logarithmic factor appears in the convergence rate. By using a block length of $l = \log n$ in the wavelet and convolution kernel estimators, the local minimax convergence rate of $(n^{-1}\log n)^{2s/(2s+1)}$ is achieved simultaneously with the global optimal rate.

**Theorem 4** *Let $\hat{f}$ be the wavelet kernel density estimator (13). Let $R$, $l$, $C_1$, $C_2$ and $\varepsilon$ be as in theorem 1, and suppose $\phi$ and $\psi$ are bounded. If $1/2 < s < N$, $\psi$ has $N-1$ vanishing moments, and*

$$c = A(0.08)^{-1}\left(C_2\|Q\|_2 + \|Q\|_1\sqrt{\frac{2N}{C_1(2N+1)}}\right)^2,$$

*then there exists a positive constant $C$ such that*

$$\sup_{f\in\Lambda^s(M,x_0,\delta)} E(\hat{f}(x_0)-f(x_0))^2 \le C(\log n/n)^{2s/(2s+1)}.$$

*Furthermore, for $l=(\log n)^{1+r}, r>0$, this upper bound is not met, i.e., local adaptivity is not attained with a block length of order larger than* $\log n$.

Since the goal is to have a single estimator that achieves both global and local minimax rates simultaneously, this theorem intentionally uses the same threshold as theorem 1. However, it could be lowered to a smaller number without a loss in the rate of convergence. This lower threshold value for local adaptivity only is

$$c > (0.08)^{-1}C_2^2\|f\|_\infty\|Q\|_2^2. \tag{17}$$

Using a block length of order larger than $\log n$, the global rate may still be attained (for example, $l=(\log n)^2$ in Hall et al. (1998)), but the local rate will not.

The threshold values in each of these theorems depends on two universal constants. The numeric value of these constants $C_1$ and $C_2$ are not specified directly in Talagrand (1994), but must be inferred from this and an earlier work of his (Talagrand (1989)). From these two papers, a value of $24e^{17/16}$ may be obtained for $C_2$. A value of $C_1$ is then derived from the relation $C_1=(C_2)^{-2}$.

The values for $\|Q\|_2$ and $\|Q\|_1$ can be obtained by examination of or numerical integration of the kernel function $K$.

To determine the value for $A$, the following method is suggested. Since the "coarse" projection $K_0 f$ does not involve the thresholding constant $c$ (recall that the linear portion of the estimator is not thresholded), use the maximum value of $\hat{K}_0(x)$ in place of $A$. Using these numeric values, the threshold $c$ for the wavelet estimator is

$$\begin{aligned} c &= \{\max_x \hat{K}_0(x)\}(0.08)^{-1}\left(24e^{17/16}\|Q\|_2+\|Q\|_1\sqrt{\frac{2N}{(24e^{17/16})^{-2}(2N+1)}}\right)^2 \\ &= \{\max_x \hat{K}_0(x)\}(24e^{17/16})^2\left(\|Q\|_2+\|Q\|_1\sqrt{2N/(2N+1)}\right)^2/(0.08). \end{aligned} \tag{18}$$

A drawback to these projection-type density estimators is that they may lead to negative values for $\hat{f}$. This problem will be overcome by using only the positive portion of the estimate. Clearly, since $f\ge 0$,

$$E\|\hat{f}_+ - f\|_2^2 \le E\|\hat{f}-f\|_2^2.$$

So the theorems are not changed by substituting $\hat{f}_+$ in for $\hat{f}$

## 4. Proofs

### *4.1 Preliminaries*

Before the proofs of the theorems are given, several preliminary results are necessary. First, a simple lemma based on Minkowski's inequality:

**Lemma 1** *Let $X_1, X_2, \dots, X_n$ be random variables. Then*

$$E\left(\sum_{i=1}^{n} X_i\right)^2 \leq \left[\sum_{i=1}^{n}(EX_i^2)^{1/2}\right]^2.$$

Second, a theorem from Talagrand (1994) as stated in Hall et al. (1998).

**Theorem 5** *Let $U_1, U_2, \dots, U_n$ be independent and identically distributed random variables. Let $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ be independent Rademacher random variables that are also independent of the $U_i$. Let $G$ be a class of functions uniformly bounded by $M$. If there exists a $v$, $H > 0$ such that for all $n$,*

$$\sup_{g \in G} \ var\ g(U) \leq v,$$

$$E \sup_{g \in G} \sum_{m=1}^{n} \varepsilon_m g(U_m) \leq nH,$$

*then there exist universal constants $C_1$ and $C_2$ such that for all $\lambda > 0$,*

$$P\left[\sup_{g \in G}\left(\frac{1}{n}\sum_{m=1}^{n} g(U_m) - Eg(U)\right) \geq \lambda + C_2 H\right] \leq e^{-nC_1\left(\frac{\lambda^2}{v} \wedge \frac{\lambda}{M}\right)}.$$

Finally, a lemma from Hall et al. (1998).

**Lemma 2** *If $K(x,y)$ is a kernel satisfying condition (1), $Q \in L^2$, and $J$ a compact interval, then*

$$E\int_J \left(\hat{K}_0(x) - K_0 f(x)\right)^2 dx \leq \|f\|_\infty \|Q\|_2^2 |J|/n,$$

*and*

$$E\int_J \left(\hat{D}_i(x) - D_i f(x)\right)^2 dx \leq 4\|f\|_\infty \|Q\|_2^2 2^i |J|/n,$$

*where $|J|$ is the length of the interval $J$.*

## *4.2 Proof of Thoerem 2*

Recall that

$$f(x) = K_0 f(x) + \sum_{i=0}^{\infty} D_i f(x)$$

and

$$\hat{f}(x) = \hat{K}_0(x) + \sum_{i=0}^{R}\sum_{k} \hat{D}_i(x) I(x \in I_{ik}) I(\hat{A}_{ik} > cn^{-1}).$$

The goal is to bound $E\|\hat{f} - f\|_2^2$. To do this, let $i_s$ be the integer such that

$$2^{i_s} \leq n^{1/(2s+1)} < 2^{i_s+1}.$$

Then Minkowksi's inequality implies that

$$
\begin{aligned}
E\|\hat{f} - f\|_2^2 \leq 4E\left\|\hat{K}_0 - K_0\right\|_2^2 &+ 4E\left\|\sum_{i=0}^{i_s}\left\{\left[\sum_k \hat{D}_i I(I_{ik}) I(\hat{A}_{ik} > cn^{-1})\right] - D_i f\right\}\right\|_2^2 \\
&+ 4E\left\|\sum_{i=i_s+1}^{R}\left\{\left[\sum_k \hat{D}_i I(I_{ik}) I(\hat{A}_{ik} > cn^{-1})\right] - D_i f\right\}\right\|_2^2 \\
&+ 4\left\|\sum_{i=R+1}^{\infty} D_i f\right\|_2^2 \\
&= T_1 + T_2 + T_3 + T_4
\end{aligned}
$$

Each piece $T_i$ will be treated individually in its own section.

*Bound on the linear part* $T_1$

To bound the linear part $T_1$, use lemma 2 and the fact that the support of $f$ is contained in $[-L, L]$.

$$
\begin{aligned}
T_1 &= 4E\|\hat{K}_0 - K_0 f\|_2^2 \\
&\leq CE\int_{-L}^{L}\left(\hat{K}_0(x) - K_0 f(x)\right)^2 dx \\
&\leq 2L\|f\|_\infty \|Q\|_2^2/n \\
&= Cn^{-1}.
\end{aligned} \tag{19}
$$

The constant $C$ is a generic constant that, for simplicity, will represent numerous constants throughout this paper.

*Bound on the nonlinear part $T_2$*

To bound the nonlinear part $T_2$, note that for a fixed $i \leq i_s$,

$$
\begin{aligned}
E\int & \left[\left(\sum_k \hat{D}_i(x) I(x \in I_{ik}) I(\hat{A}_{ik} > cn^{-1})\right) - D_i f(x)\right]^2 dx \\
&\leq E\sum_k \int_{I_{ik}} \left(\hat{D}_i(x) - D_i f(x)\right)^2 dx\ I(\hat{A}_{ik} > cn^{-1}) \\
&\quad + E\sum_k \int_{I_{ik}} (D_i f(x))^2\, dx\ I(\hat{A}_{ik} \leq cn^{-1}) I(A_{ik} \leq 2cn^{-1}) \\
&\quad + E\sum_k \int_{I_{ik}} (D_i f(x))^2\, dx\ I(\hat{A}_{ik} \leq cn^{-1}) I(A_{ik} > 2cn^{-1}) \\
&\leq E\int \left(\hat{D}_i(x) - D_i f(x)\right)^2 dx \\
&\quad + E\sum_k \int_{I_{ik}} (D_i f(x))^2\, dx\ I(A_{ik} \leq 2cn^{-1}) \\
&\quad + E\sum_k \int_{I_{ik}} (D_i f(x))^2\, dx\ I(\hat{A}_{ik} \leq cn^{-1}) I(A_{ik} > 2cn^{-1}) \\
&= T_{21} + T_{22} + T_{23}.
\end{aligned}
$$

$T_{21}$ and $T_{22}$ are bounded as in Hall, et al. Hall et al. (1998):

$$
T_{21}, T_{22} \leq C2^i/n. \tag{20}
$$

To bound $T_{23}$, the following lemma from Hall et al. (1998) is useful:

**Lemma 3** *If $\int_{I_{ik}} (D_i f(x))^2\, dx \leq lc/(2n)$ then*

$$
\left\{\int_{I_{ik}} \left(\hat{D}_i(x)\right)^2 dx \geq lc/n\right\} \subseteq \left\{\int_{I_{ik}} \left(\hat{D}_i(x) - D_i f(x)\right)^2 dx \geq 0.08lc/n\right\},
$$

*and if $\int_{I_{ik}} (D_i f(x))^2\, dx > 2lc/n$ then*

$$
\left\{\int_{I_{ik}} \left(\hat{D}_i(x)\right)^2 dx \leq lc/n\right\} \subseteq \left\{\int_{I_{ik}} \left(\hat{D}_i(x) - D_i f(x)\right)^2 dx \geq 0.16lc/n\right\}.
$$

Using this lemma,

$$
T_{23} \leq E\sum_k \int_{I_{ik}} (D_i f(x))^2\, dx\ I\left(\int_{I_{ik}} \left(\hat{D}_i(x) - D_i f(x)\right)^2 dx \geq 0.16lc/n\right).
$$

By (8), and the fact that the length of the interval $I_{ik}$ is $l/2^i$,

$$\begin{aligned}\int_{I_{ik}} (D_i f(x))^2\,dx &= \int_{I_{ik}} (K_{i+1}f(x) - K_i f(x))^2 dx \\ &\le 2\int_{I_{ik}} (K_{i+1}f(x))^2 dx + 2\int_{I_{ik}} (K_i f(x))^2 dx \\ &\le 2\int_{I_{ik}} \|K_{i+1}f\|_\infty^2 dx + 2\int_{I_{ik}} \|K_i f\|_\infty^2 dx \\ &\le 4\|f\|_\infty^2 \|Q\|_1^2 l/2^i.\end{aligned}$$

So,

$$T_{23} \le 4\|f\|_\infty^2\|Q\|_1^2 l/2^i \sum_k P\left(\left[\int_{I_{ik}} \left(\hat{D}_i(x) - D_i f(x)\right)^2 dx\right]^{1/2} \ge \sqrt{0.16lc/n}\right). \quad (21)$$

To bound the above probability, Talagrand's theorem (theorem 5) will be used. From Hall, et al. it is shown that

$$\begin{aligned}&\left\{\int_{I_{ik}} \left(\hat{D}_i(x) - D_i f(x)\right)^2 dx\right\}^{1/2} \\ &\qquad = \sup_{g\in G}\left\{\frac{1}{n}\sum_{m=1}^{n}\int_{I_{ik}} g(x)D_i(x, X_m)dx - E\int_{I_{ik}} g(x)D_i(x, X_1)dx\right\},\end{aligned}$$

where the function set $G$ is

$$G = \left\{\int_{I_{ik}} g(x)D_i(x,\cdot)dx : \|g\|_2 \le 1\right\},$$

and values for $M$, $v$, and $H$ in theorem 5 are:

$$M = 2^{i/2}\|Q\|_2,$$

$$v = \|f\|_\infty\|Q\|_1^2,$$

and

$$H = \sqrt{l\|f\|_\infty\|Q\|_2^2/n}.$$

Letting $\lambda = \sqrt{0.16lc/n} - C_2\sqrt{l\|f\|_\infty\|Q\|_2^2/n} > 0$, Talagrand's theorem then implies

$$\begin{aligned}&P\left(\left[\int_{I_{ik}} \left(\hat{D}_i(x) - D_i f(x)\right)^2 dx\right]^{1/2} \ge \lambda + C_2\sqrt{l\|f\|_\infty\|Q\|_2^2/n}\right) \\ &\qquad \le \exp\left\{-nC_1\left[\left(\lambda^2/\|f\|_\infty\|Q\|_1^2\right) \wedge \left(\lambda/(2^{i/2}\|Q\|_2)\right)\right]\right\}.\end{aligned}$$

Now, if $0 \le i \le i_s$, then $\lambda^2/\left(\|f\|_\infty \|Q\|_1^2\right) < \lambda/\left(2^{i/2}\|Q\|_2\right)$ for large $n$ and positive $\lambda$:

$$\begin{aligned}
&\lambda^2/\left(\|f\|_\infty \ \|Q\|_1^2\right) \le \lambda/\left(2^{i/2}\|Q\|_2\right) \\
&\qquad \Leftrightarrow \lambda \le \|f\|_\infty \|Q\|_1^2/\left(2^{i/2}\|Q\|_2\right) \\
&\qquad \Leftrightarrow \sqrt{0.16lc/n} - C_2\sqrt{l\|f\|_\infty \|Q\|_2^2/n} \le \frac{\|f\|_\infty \|Q\|_1^2}{\|Q\|_2} 2^{-i/2} \\
&\qquad \Leftarrow \sqrt{l/n}\left(\sqrt{0.16c} - C_2\sqrt{\|f\|_\infty \|Q\|_2^2}\right) \le \frac{\|f\|_\infty \|Q\|_1^2}{\|Q\|_2} 2^{-i_s/2} \qquad (22) \\
&\qquad \Leftrightarrow \sqrt{2^{i_s}\log n/n} \le \frac{\|f\|_\infty \|Q\|_1^2}{\left(\sqrt{0.16c} - C_2\sqrt{\|f\|_\infty \|Q\|_2^2}\right)\|Q\|_2} \\
&\qquad \Leftrightarrow n^{-2s/(2s+1)}\log n \le \frac{\|f\|_\infty^2 \|Q\|_1^4}{\left(\sqrt{0.16c} - C_2\sqrt{\|f\|_\infty \|Q\|_2^2}\right)^2 \|Q\|_2^2} \\
&\qquad \Leftarrow n \text{ large enough, say } n \ge n^*.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&P\left(\left[\int_{I_{ik}} \left(\hat{D}_i(x) - D_i f(x)\right)^2 dx\right]^{1/2} \ge \sqrt{0.16lc/n}\right) \\
&\quad \le C \cdot \exp\left[-nC_1\lambda^2/\|f\|_\infty \|Q\|_1^2\right] \\
&\quad = C \cdot \exp\left[-nC_1\left(\sqrt{0.16lc/n} - C_2\sqrt{l\|f\|_\infty \|Q\|_2^2/n}\right)^2 /\|f\|_\infty \|Q\|_1^2\right] \qquad (23) \\
&\quad = C \cdot \exp\left[-C_1\left(\sqrt{0.16c} - C_2\sqrt{\|f\|_\infty \|Q\|_2^2}\right)^2 \|f\|_\infty^{-1}\|Q\|_1^{-2}\log n\right] \\
&\quad = Cn^{-\delta},
\end{aligned}$$

where $\delta$ is the constant

$$\delta = \frac{C_1\left(\sqrt{0.16c} - C_2\sqrt{\|f\|_\infty \|Q\|_2^2}\right)^2}{\|f\|_\infty \|Q\|_1^2}.$$

Putting (21) and (23) together with the fact that the number of nonoverlapping intervals $I_{ik}$ that intersect the support of $f$ is no more than $2L2^i/l$,

$$T_{23} \le Cn^{-\delta}. \qquad (24)$$

All pieces are now available to bound $T_2$. Using Minkowksi's inequality,

$$\begin{aligned}
T_2 &= 4E\left\|\sum_{i=0}^{i_s}\left\{\left(\sum_k \hat{D}_i I(I_{ik}) I(\hat{A}_{ik} > cn^{-1})\right) - D_i f\right\}\right\|_2^2 \\
&\le 4E\left(\sum_{i=0}^{i_s}\left\|\left\{\left(\sum_k \hat{D}_i I(I_{ik}) I(\hat{A}_{ik} > cn^{-1})\right) - D_i f\right\}\right\|_2\right)^2,
\end{aligned}$$

and using lemma 1 with $X_i = \left\| \left\{ \left( \sum_k \hat{D}_i I(I_{ik}) I(\hat{A}_{ik} > cn^{-1}) \right) - D_i f \right\} \right\|_2$,

$$
\begin{aligned}
T_2 &\le 4 \left[ \sum_{i=0}^{i_s} \left( E \left\| \left[ \left( \sum_k \hat{D}_i I(I_{ik}) I(\hat{A}_{ik} > cn^{-1}) \right) - D_i f \right] \right\|_2^2 \right)^{1/2} \right]^2 \\
&= 4 \left[ \sum_{i=0}^{i_s} \left( E \int \left[ \left( \sum_k \hat{D}_i(x) I(x \in I_{ik}) I(\hat{A}_{ik} > cn^{-1}) \right) - D_i f(x) \right]^2 dx \right)^{1/2} \right]^2 \\
&\le 4 \left( \sum_{i=0}^{i_s} (T_{21} + T_{22} + T_{23})^{1/2} \right)^2 \\
&\le C \left( \sum_{i=0}^{i_s} T_{21}^{1/2} + T_{22}^{1/2} + T_{23}^{1/2} \right)^2 .
\end{aligned}
$$

Using the bounds from (20) and (24),

$$
\begin{aligned}
T_2 &\le C \left[ \sum_{i=0}^{i_s} \left( \left( 2^i/n \right)^{1/2} + n^{-\delta/2} \right) \right]^2 \\
&\le C \left( 2^{i_s} n^{-1} + i_s^2 n^{-\delta} \right) \\
&= C \left[ n^{-2s/(2s+1)} + (\log_2 n^{1/(2s+1)})^2 n^{-\delta} \right] .
\end{aligned} \tag{25}
$$

*Bound on the nonlinear part $T_3$*

For a fixed $i, i_s + 1 \le i \le R$,

$$
\begin{aligned}
E \int & \left[ \left( \sum_k \hat{D}_i(x) I(x \in I_{ik}) I(\hat{A}_{ik} > cn^{-1}) \right) - D_i f(x) \right]^2 dx \\
&\le E \sum_k \int_{I_{ik}} \left( \hat{D}_i(x) - D_i f(x) \right)^2 dx \; I(\hat{A}_{ik} > cn^{-1}) I(A_{ik} > c/(2n)) \\
&\quad + E \sum_k \int_{I_{ik}} \left( \hat{D}_i(x) - D_i f(x) \right)^2 dx \; I(\hat{A}_{ik} > cn^{-1}) I(A_{ik} \le c/(2n)) \\
&\quad + E \sum_k \int_{I_{ik}} (D_i f(x))^2 \, dx \; I(\hat{A}_{ik} \le cn^{-1}) I(A_{ik} \le 2cn^{-1}) \\
&\quad + E \sum_k \int_{I_{ik}} (D_i f(x))^2 \, dx \; I(\hat{A}_{ik} \le cn^{-1}) I(A_{ik} > 2cn^{-1}) \\
&= T_{31} + T_{32} + T_{33} + T_{34}.
\end{aligned}
$$

$T_{31}$ and $T_{33}$ are treated the same as in Hall, et al. Hall et al. (1998) and have the same bounds to within a constant:

$$
T_{31}, T_{33} \le C \left( 2^{-2is} + la(n) n^{1/(2N+1)} n^{r-1} 2^{-ir} \right) , \tag{26}
$$

where $r > 0$ is arbitrary and $a(n) = O(n^\zeta)$ for all $\zeta > 0$.

By lemma 3,

$$\begin{aligned} T_{32} &= \sum_k E \int_{I_{ik}} \left(\hat{D}_i(x) - D_i f(x)\right)^2 dx \; I(\hat{A}_{ik} > cn^{-1}) I(A_{ik} \le c/(2n)) \\ &\le \sum_k E \left[ \int_{I_{ik}} \left(\hat{D}_i(x) - D_i f(x)\right)^2 dx \right. \\ &\quad \left. \cdot I \left( \left\{ \int_{I_{ik}} \left(\hat{D}_i(x) - D_i f(x)\right)^2 dx \right\}^{1/2} \ge \sqrt{0.08lc/n} \right) \right]. \end{aligned}$$

To bound this, note that for any non-negative random variable $Y$,

$$\begin{aligned} EY^2 I(Y > a) &= \int_0^\infty P\left(Y^2 I(Y > a) > u\right) du \\ &= \int_0^{a^2} P\left(Y^2 I(Y > a) > a^2\right) du + \int_{a^2}^\infty P\left(Y^2 I(Y > a) > u\right) du \\ &= \int_0^{a^2} P\left(Y > a\right) du + \int_{a^2}^\infty P\left(Y^2 > u\right) du \\ &= a^2 P(Y > a) + \int_a^\infty 2y P(Y > y) dy, \end{aligned}$$

The integrals in $T_{32}$ are of this form with

$$Y = \left[ \int_{I_{ik}} \left(\hat{D}_i(x) - D_i f(x)\right)^2 dx \right]^{1/2} \ge 0$$

and $a = \sqrt{0.08lc/n}$. From Talagrand's theorem,

$$\begin{aligned} P(Y > y) &= P[Y > (y - C_2 H) + C_2 H] \\ &\le \exp\left[ -nC_1 \left( \frac{(y - C_2 H)^2}{\|f\|_\infty \|Q\|_1^2} \wedge \frac{y - C_2 H}{2^{i/2} \|Q\|_2} \right) \right] \end{aligned}$$

and therefore

$$\begin{aligned} EY^2 I(Y > a) &\le a^2 \exp\left[ -nC_1 \left( \frac{(a - C_2 H)^2}{\|f\|_\infty \|Q\|_1^2} \wedge \frac{a - C_2 H}{2^{i/2} \|Q\|_2} \right) \right] \\ &\quad + \int_a^\infty 2y \exp\left[ -nC_1 \left( \frac{(y - C_2 H)^2}{\|f\|_\infty \|Q\|_1^2} \wedge \frac{y - C_2 H}{2^{i/2} \|Q\|_2} \right) dy \right] \\ &= T_{321} + T_{322}. \end{aligned}$$

Now, $(a - C_2H)^2 \|f\|_\infty^{-1}\|Q\|_1^{-2} \le (a - C_2H)2^{-i/2}\|Q\|_2^{-1}$ for large $n$ and $(a - C_2H)$ positive:

$$(a - C_2H)^2/ \left(\|f\|_\infty \|Q\|_1^2\right) \le (a - C_2H)/ \left(2^{i/2}\|Q\|_2\right)$$

$$\Leftarrow \sqrt{l/n}\left(\sqrt{0.08c} - C_2\sqrt{\|f\|_\infty\|Q\|_2^2}\right) \le \frac{\|f\|_\infty\|Q\|_1^2}{\|Q\|_2} 2^{-R/2} \tag{27}$$

$$\Leftrightarrow 2^R \log n/n \le \frac{\|f\|_\infty^2\|Q\|_1^4}{\left(\sqrt{0.08c} - C_2\sqrt{\|f\|_\infty\|Q\|_2^2}\right)^2 \|Q\|_2^2}$$

$$\Leftarrow 2^R \le n^{1-\varepsilon} \text{ for some fixed } \varepsilon > 0 \text{ and } n \ge n'.$$

Therefore,

$$\begin{aligned} T_{321} &\le C\left(\sqrt{0.08cl/n}\right)^2 \exp\left\{-nC_1\left[\frac{\left(\sqrt{0.08cl/n} - C_2\sqrt{\|f\|_\infty\|Q\|_2^2 l/n}\right)^2}{\|f\|_\infty\|Q\|_1^2}\right]\right\} \\ &= Cn^{-1}\log n \exp\left\{-C_1\left[\frac{\left(\sqrt{0.08c} - C_2\sqrt{\|f\|_\infty\|Q\|_2^2}\right)^2}{\|f\|_\infty\|Q\|_1^2}\right]\log n\right\} \\ &\le Cn^{-\gamma-1}\log n, \end{aligned} \tag{28}$$

where $\gamma$ is the constant

$$\gamma = \frac{C_1\left(\sqrt{0.08c} - C_2\sqrt{\|f\|_\infty\|Q\|_2^2}\right)^2}{\|f\|_\infty\|Q\|_1^2}. \tag{29}$$

For $T_{322}$, first assume that

$$a \le a_0 = \frac{\|f\|_\infty\|Q\|_1^2}{\|Q\|_2 2^{i/2}} + C_2\sqrt{l\|f\|_\infty\|Q\|_2/n}.$$

Then,

$$\begin{aligned} T_{322} &= \int_a^{a_0} 2y \exp\left(-nC_1\frac{(y - C_2H)^2}{\|f\|_\infty\|Q\|_1^2}\right) dy + \int_{a_0}^{\infty} 2y \exp\left(-nC_1\frac{y - C_2H}{2^{i/2}\|Q\|_2}\right) dy \\ &= T_{3221} + T_{3222}. \end{aligned}$$

To bound $T_{3221}$, note that by a change of variables and increase in upper limit of integration,

$$\begin{aligned} T_{3221} &\le \int_{a-C_2H}^{\infty} 2y \exp\left(-nC_1\frac{y^2}{\|f\|_\infty\|Q\|_1^2}\right) dy \\ &\quad + \int_{a-C_2H}^{\infty} 2C_2H \exp\left(-nC_1\frac{y^2}{\|f\|_\infty\|Q\|_1^2}\right) dy \\ &= \|f\|_\infty\|Q\|_1^2 n^{-1}C_1^{-1} \exp\left(-nC_1\frac{(a - C_2H)^2}{\|f\|_\infty\|Q\|_1^2}\right) \\ &\quad + \int_{a-C_2H}^{\infty} 2C_2Hy(1/y) \exp\left(-nC_1\frac{y^2}{\|f\|_\infty\|Q\|_1^2}\right) dy. \end{aligned} \tag{30}$$

The first term on the right of (30) is bounded by $Cn^{-1-\gamma}$, where $\gamma$ is the constant in (29). To bound the the second term, use integration by parts.

$$\begin{aligned}\int_{a-C_2H}^{\infty} & 2C_2Hy(1/y)\exp\left(-nC_1\frac{y^2}{\|f\|_\infty\|Q\|_1^2}\right)dy \\ &= \frac{C_2H}{a-C_2H}\frac{\|f\|_\infty\|Q\|_1^2}{nC_1}\exp\left(-nC_1\frac{(a-C_2H)^2}{\|f\|_\infty\|Q\|_1^2}\right) \\ &\quad -\int_{a-C_2H}^{\infty} C_2H\frac{\|f\|_\infty\|Q\|_1^2}{nC_1}\frac{1}{y^2}\exp\left(-nC_1\frac{y^2}{\|f\|_\infty\|Q\|_1^2}\right)dy.\end{aligned}$$

Since the integrand in second term on the right side above is strictly positive,

$$\begin{aligned}\int_{a-C_2H}^{\infty} & 2C_2Hy(1/y)\exp\left(-nC_1\frac{y^2}{\|f\|_\infty\|Q\|_1^2}\right)dy \\ &\le \frac{C_2\sqrt{\|f\|_\infty\|Q\|_2^2\log n/n}}{\sqrt{0.08c\log n/n}-C_2\sqrt{\|f\|_\infty\|Q\|_2^2\log n/n}}\frac{\|f\|_\infty\|Q\|_1^2}{nC_1}\exp\left(-\gamma\log n\right) \\ &\le Cn^{-1-\gamma},\end{aligned}$$

where again $\gamma$ is from (29).

Now to bound $T_{3222}$. Using integration by parts,

$$\begin{aligned}T_{3222} &= \int_{a_0}^{\infty} 2y\exp\left(-nC_1\frac{y-C_2H}{2^{i/2}\|Q\|_2}\right)dy \\ &= 2a_0\frac{2^{i/2}\|Q\|_2}{nC_1}\exp\left(-nC_1\frac{a_0-C_2H}{2^{i/2}\|Q\|_2}\right) \\ &\quad +\int_{a_0}^{\infty} 2\frac{2^{i/2}\|Q\|_2}{nC_1}\exp\left(-nC_1\frac{y-C_2H}{2^{i/2}\|Q\|_2}\right)dy.\end{aligned} \tag{31}$$

The first term of (31) is bounded above by

$$\begin{aligned}Cn^{-1}2^{i/2} & \left(\frac{\|f\|_\infty\|Q\|_1^2}{\|Q\|_2 2^{i/2}}+C_2\sqrt{l\|f\|_\infty\|Q\|_2/n}\right)\exp\left(-\frac{nC_1\|f\|_\infty\|Q\|_1^2}{2^i\|Q\|_2^2}\right) \\ &\le C\left(n^{-1}+n^{-1}\sqrt{2^i\log n/n}\right)\exp\left(-\frac{nd}{2^i}\right),\end{aligned}$$

and the second term is no more than

$$C\left(n^{-1}2^{i/2}\right)^2\exp\left(-\frac{nC_1\|f\|_\infty\|Q\|_1^2}{2^i\|Q\|_2^2}\right)\le C2^in^{-2}\exp\left(-\frac{nd}{2^i}\right)$$

where $d$ is the constant

$$d=\frac{C_1\|f\|_\infty\|Q\|_1^2}{\|Q\|_2^2}. \tag{32}$$

If $a>a_0$, then,

$$T_{3222}=\int_a^{a_0} 2y\exp\left(-nC_1\frac{(y-C_2H)^2}{\|f\|_\infty\|Q\|_1^2}\right)dy+\int_{a_0}^{\infty} 2y\exp\left(-nC_1\frac{y-C_2H}{2^{i/2}\|Q\|_2}\right)dy$$

Since the first integral is strictly negative, the same bound holds as shown above in (31). Therefore,

$$T_{32} = \sum_k (T_{321} + T_{3221} + T_{3222})$$
$$\leq C \sum_k \left[ n^{-\gamma-1} \log n + n^{-\gamma-1} + \left( n^{-1} + n^{-1}\sqrt{2^i \log n / n} + 2^i n^{-2} \right) \exp\left( -\frac{nd}{2^i} \right) \right]. \tag{33}$$

Since the number of blocks $k$ intersecting the support of $f$ is no more than $2L2^i/\log n$,

$$T_{32} \leq C2^i/\log n \left[ n^{-\gamma-1} \log n + \left( n^{-1} + n^{-1}\sqrt{2^i \log n/n} + 2^i n^{-2} \right) \exp\left( -\frac{nd}{2^i} \right) \right]. \tag{34}$$

To bound the piece $T_{32}$ using Talagrand's theorem, it is critical that $2^R \leq n^{1-\varepsilon}$ (or, more generally, $2^R \leq Cn^{1-\varepsilon}$). If $2^R \geq n$, then for $i = R$, the argument at (27) is invalid. Letting $i = R$,

$$\frac{(a - C_2 H)^2}{\|f\|_\infty \|Q\|_1^2} = Cn^{-1} \log n,$$

which, for large $n$, is greater than

$$\frac{a - C_2 H}{2^{i/2} \|Q\|_2} = Cn^{-1} \sqrt{\log n}.$$

Thus, at (28) the bound becomes (for $i = R$)

$$T_{321} \leq C \left( \sqrt{0.08cl/n} \right)^2 \exp\left[ -nC_1 \left( \frac{\sqrt{0.08cl/n} - C_2 \sqrt{\|f\|_\infty \|Q\|_2^2 l/n}}{n^{1/2} \|Q\|_2} \right) \right]$$
$$= Cn^{-1} \log n \exp\left( -D\sqrt{\log n} \right)$$

for some constant $D$. This equates to adding the term $\exp\left(-D\sqrt{\log n}\right)$ to the bound at (34). Now, for any constant $C$ and all $s$ in $(1/2, N - \rho)$ there exists $n$ large enough such that

$$C \cdot \frac{2s}{2s+1} > \frac{D}{\sqrt{\log n}} \Rightarrow C \cdot \frac{2s}{2s+1} \log n > D\sqrt{\log n}$$
$$\Rightarrow Cn^{\frac{2s}{2s+1}} > \exp\left( D\sqrt{\log n} \right)$$
$$\Rightarrow Cn^{-\frac{2s}{2s+1}} < \exp\left( -D\sqrt{\log n} \right).$$

This added term therefore prevents the estimator from attaining the optimal minimax rate of convergence.

The only way around this problem without finding a sharper bound than that provided by theorem 5 is if the block length $l$ were $(\log n)^2$ or larger. However, as pointed out in section 3, this block length will not result in optimal local adaptivity.

Also, for (27) to remain true for values of $i$ approaching $R$, larger and larger values of the constant $C$ are necessary in (28). No single constant $C$ would suffice in (28).

To bound $T_{34}$, observe that the only difference between $T_{23}$ and $T_{34}$ is the range of the index $i$. Therefore, by repeating the argument for $T_{23}$, the bound for $T_{34}$ is the same as at (24):

$$T_{34} \le Cn^{-\delta}. \tag{35}$$

To bound $T_3$, use lemma 5 and Minkowksi's inequality in a manner similar to the treatment of $T_2$.

$$\begin{aligned}
T_3 &= 4E\left\| \sum_{i=i_s+1}^{R} \left[ \left( \sum_k \hat{D}_i I(I_{ik}) I(\hat{A}_{ik} > cn^{-1}) \right) - D_i f \right] \right\|_2^2 \\
&\le 4\left[ \sum_{i=i_s+1}^{R} \left( E \int \left[ \left( \sum_k \hat{D}_i(x) I(x \in I_{ik}) I(\hat{A}_{ik} > cn^{-1}) \right) - D_i f(x) \right]^2 dx \right)^{1/2} \right]^2 \\
&\le C \left( \sum_{i=i_s+1}^{R} T_{31}^{1/2} + T_{32}^{1/2} + T_{33}^{1/2} + T_{34}^{1/2} \right)^2 \qquad (36) \\
&\le C \left[ \left( \sum_{i=i_s+1}^{R} T_{31}^{1/2} \right)^2 + \left( \sum_{i=i_s+1}^{R} T_{32}^{1/2} \right)^2 + \left( \sum_{i=i_s+1}^{R} T_{33}^{1/2} \right)^2 + \left( \sum_{i=i_s+1}^{R} T_{34}^{1/2} \right)^2 \right].
\end{aligned}$$

First, for $j = 1$ or 3, from (26) we have

$$\begin{aligned}
\left( \sum_{i=i_s+1}^{R} T_{3j}^{1/2} \right)^2 &\le C \left[ \sum_{i=i_s+1}^{R} \left( 2^{-2is} + (\log n) n^{1/(2N+1)+\zeta} n^{r-1} 2^{-ir} \right)^{1/2} \right]^2 \\
&\le C \left( \sum_{i=i_s+1}^{R} 2^{-is} + \sqrt{(\log n) n^{1/(2N+1)+\zeta} n^{r-1} 2^{-ir}} \right)^2 \\
&\le C \left[ n^{-2s/(2s+1)} + (\log n)\, n^{1/(2N+1)+\zeta+r-1} \left( \sum_{i=i_s+1}^{R} 2^{-ir/2} \right)^2 \right] \\
&\le C \left[ n^{-2s/(2s+1)} + (\log n)\, n^{1/(2N+1)+\zeta+r-1-r/(2s+1)} \right].
\end{aligned}$$

Now,

$$n^{1/(2N+1)+\zeta+r-1-r/(2s+1)} < n^{-2s/(2s+1)}$$

whenever

$$1/(2s+1) - 1/(2N+1) > 2rs/(2s+1) + \zeta. \tag{37}$$

Since $s < N$ and $\zeta, r > 0$ are arbitrary, this is easily accomplished. Let $\kappa$ be the positive integer such that for the /zeta and $r$ making (37) true,

$$n^{\kappa} n^{1/(2N+1)+\zeta+r-1-r/(2s+1)} = n^{-2s/(2s+1)}.$$

Then,

$$(\log n)\, n^{1/(2N+1)+\zeta+r-1-r/(2s+1)} = (\log n)\, n^{-\kappa} n^{-2s/(2s+1)}$$
$$\leq C n^{-2s/(2s+1)},$$

and

$$\left( \sum_{i=i_s+1}^{R} T_{3j}^{1/2} \right)^2 \leq C \left( n^{-2s/(2s+1)} \right). \tag{38}$$

Next, from (34),

$$\left( \sum_{i=i_s+1}^{R} T_{32}^{1/2} \right)^2 \leq C \left[ \left( \sum_{i_s+1}^{R} \left(2^i n^{-(\gamma+1)}\right)^{1/2} \right)^2 + \left( \sum_{i_s+1}^{R} \left( 2^i (n \log n)^{-1} \sqrt{2^i \log n / n} e^{-\frac{nd}{2^i}} \right)^{1/2} \right)^2 + \left( \sum_{i_s+1}^{R} \left( 2^{2i} (n^2 \log n)^{-1} e^{-\frac{nd}{2^i}} \right)^{1/2} \right)^2 + \left( \sum_{i_s+1}^{R} \left( 2^i (n \log n)^{-1} e^{-\frac{nd}{2^i}} \right)^{1/2} \right)^2 \right].$$

Observe that for large $n$ and positive $d$,

$$\log n \leq n^{\varepsilon} d \Rightarrow n \leq e^{n^{\varepsilon} d} = e^{\frac{nd}{n^{1-\varepsilon}}}.$$

Therefore,

$$2^i \leq n \leq e^{\frac{nd}{2^i}} \text{ for all } i = 0, 1, \ldots, R.$$

So,

$$2^i e^{-\frac{nd}{2^i}} \leq C \text{ for large } n.$$

Therefore,

$$\begin{aligned}
\left(\sum_{i=i_s+1}^{R} T_{32}^{1/2}\right)^2 &\le C\left[n^{-(\gamma+1)}\left(\sum_{i_s+1}^{R} 2^{i/2}\right)^2 + \left(\sum_{i_s+1}^{R}\left(n^{-1}\sqrt{2^i/(n\log n)}\right)^{1/2}\right)^2\right. \\
&\quad \left. + \left(\sum_{i_s+1}^{R}\left(2^i(n^2\log n)^{-1}\right)^{1/2}\right)^2 + \left(\sum_{i_s+1}^{R}(n\log n)^{-1/2}\right)^2\right] \\
&\le C\left[2^R n^{-(\gamma+1)} + \left(n\sqrt{n\log n}\right)^{-1}\left(\sum_{i_s+1}^{R} 2^{i/4}\right)^2\right. \\
&\quad \left. + \left(n^2\log n\right)^{-1}\left(\sum_{i_s+1}^{R} 2^{i/2}\right)^2 + (n\log n)^{-1} R^2\right] \qquad (39) \\
&\le C\left[n^{-\gamma} + \left(n\sqrt{n\log n}\right)^{-1} 2^{R/2} + \left(n^2\log n\right)^{-1} 2^R\right. \\
&\quad \left. + (n\log n)^{-1}\left(\log_2 n^{1-\varepsilon}\right)^2\right] \\
&\le C\left(n^{-\gamma} + n^{-1}\right).
\end{aligned}$$

And finally, from (35),

$$\left(\sum_{i=i_s+1}^{R} T_{34}^{1/2}\right)^2 \le \left(\log_2 n^{1-\varepsilon}\right)^2 n^{-\delta}. \qquad (40)$$

Putting (38), (39), (40) together yields

$$T_3 \le C\left[n^{-2s/(2s+1)} + n^{-\gamma} + \left(\log_2 n^{1-\varepsilon}\right)^2 n^{-\delta}\right]. \qquad (41)$$

*Bound on the nonlinear part $T_4$*

From Hall et al. (1998),

$$||D_i f||_2^2 \le C\left(2^{-2si} + a(n)n^{1/(2N+1)}2^{-i}\right),$$

where $a(n) = O(n^{\zeta})$ for all $\zeta > 0$. Then, by Minkowski's inequality,

$$\begin{aligned} T_4 &= 4\left\|\sum_{i=R+1}^{\infty} D_i f\right\|_2^2 \\ &\le C\left(\sum_{i=R+1}^{\infty} ||D_i f||_2\right)^2 \\ &\le C\left[\sum_{i=R+1}^{\infty}\left(2^{-2si} + a(n)n^{1/(2N+1)}2^{-i}\right)^{1/2}\right]^2 \\ &\le C\left[\left(\sum_{i=R+1}^{\infty} 2^{-si}\right)^2 + a(n)n^{1/(2N+1)}\left(\sum_{i=R+1}^{\infty} 2^{-i/2}\right)^2\right] \\ &= C\left(2^{-2Rs} + a(n)n^{1/(2N+1)}2^{-R}\right) \\ &= C\left(n^{-2s(1-\varepsilon)} + n^{\zeta+1/(2N+1)-1+\varepsilon}\right). \end{aligned} \tag{42}$$

*Determination of constants $\gamma, \delta$, $\varepsilon$, and $c$*

Using the bounds from (19), (25), (41), and (42),

$$\begin{aligned} E||f-\hat{f}||_2^2 \le C\Big[&n^{-2s/(2s+1)} + \left(\log_2 n^{1/(2s+1)}\right)^2 n^{-\delta} + n^{-\gamma} \\ &+ \left(\log_2 n^{1-\varepsilon}\right)^2 n^{-\delta} + n^{-2s(1-\varepsilon)} + n^{\zeta+1/(2N+1)-1+\varepsilon}\Big]. \end{aligned}$$

First, a bound for $\varepsilon$. Note that

$$n^{\zeta+1/(2N+1)-1+\varepsilon} \le n^{-2s/(2s+1)} \tag{43}$$

if and only if

$$\frac{2N}{2N+1} - \frac{2s}{2s+1} \ge \zeta + \varepsilon.$$

Since it is desired that (43) hold for all $1/2 < s < N-\rho$, choose $\varepsilon$ and $\zeta$ such that

$$\zeta + \varepsilon \le \frac{2\rho}{(2N+1)(2(N-\rho)+1)}.$$

$\zeta$ can be any arbitrary positive number, so for simplicity take it to be the same as $\varepsilon$. Then (43) is satisfied if

$$\varepsilon = \frac{\rho}{(2N+1)(2(N-\rho)+1)}. \tag{44}$$

This choice of $\varepsilon$ is less than $1/2$, so

$$n^{-2s(1-\varepsilon)} \leq n^{-s} \leq n^{-2s/(2s+1)},$$

for all $1/2 < s < N - \rho$.

For $\gamma$,

$$n^{-\gamma} \leq n^{-2s/(2s+1)}$$

for all $1/2 < s < N - \rho$ if and only if

$$\gamma = \frac{C_1\left(\sqrt{0.08c} - C_2\sqrt{\|f\|_\infty\|Q\|_2^2}\right)^2}{\|f\|_\infty\|Q\|_1^2} \geq \frac{2(N-\rho)}{2(N-\rho)+1}.$$

The above constraint is met if the value of the threshold $c$ is set accordingly:

$$c \geq (0.08)^{-1}\left(C_2\sqrt{\|f\|_\infty\|Q\|_2^2} + \sqrt{\frac{\|f\|_\infty\|Q\|_1^2 2(N-\rho))}{C_1(2(N-\rho)+1)}}\right)^2,$$

or, since $\|f\|_\infty$ is unknown but bounded by $A$,

$$c = A(0.08)^{-1}\left(C_2\|Q\|_2 + \|Q\|_1\sqrt{\frac{2(N-\rho))}{C_1(2(N-\rho)+1)}}\right)^2. \tag{45}$$

Note that the condition at (27) that $a - C_2H$ be positive is met if (45) holds.

$$\begin{aligned} a - C_2H &= \sqrt{0.08cn^{-1}\log n} - C_2\sqrt{n^{-1}\log n\|f\|_\infty\|Q\|_2^2} \\ &= \sqrt{n^{-1}\log n}\left(\sqrt{0.08c} - C_2\sqrt{\|f\|_\infty\|Q\|_2^2}\right) \\ &> \sqrt{n^{-1}\log n}\left(\sqrt{AC_2^2\|Q\|_2^2} - C_2\sqrt{\|f\|_\infty\|Q\|_2^2}\right) \\ &= \sqrt{n^{-1}\log n}C_2\|Q\|_2\left(\sqrt{A} - \sqrt{\|f\|_\infty}\right) \\ &\geq 0. \end{aligned}$$

For $\delta$, note that

$$\delta = \frac{C_1\left(\sqrt{0.16c} - C_2\sqrt{\|f\|_\infty\|Q\|_2^2}\right)^2}{\|f\|_\infty\|Q\|_1^2} > \gamma. \tag{46}$$

Therefore

$$\left[\left(\log_2 n^{1/(2s+1)}\right)^2 + \left(\log_2 n^{1-\varepsilon}\right)^2\right] n^{-\delta} < Cn^{-2s/(2s+1)}.$$

A similar argument to the one above shows the condition at (22) that $\lambda$ be positive is met if (46) holds.

Therefore, using the bounds for $\varepsilon$ and $c$ at (44) and (45),

$$E\|f-\hat{f}\|_2^2 \le Cn^{-2s/(2s+1)},$$

and the theorem is proved.

One might be tempted to conclude that since the upper bound is $N-\rho$ for any $\rho>0$, the upper bound for the unknown smoothness parameter $s$ is really $N$. However, by letting $\rho$ go to 0, (44) shows that $\varepsilon$ goes to 0 and the argument for the term $T_{321}$ at (27) is not valid. See the discussion in section 4.2.3 following the bound for $T_{32}$.

## *4.3 Proof of Theorem 3*

The proof of theorem 3 follows closely that of theorem 2. Substitute $\zeta$ for $\varepsilon$ throughout the proof of theorem 2. The only remaining differences are those involving the pieces $T_{31}$, $T_{33}$, and $T_4$. These are the pieces that involve the number of discontinuities.

First, from (26)

$$T_{31}, T_{33} \le D\left(2^{-2is} + (\log n) n^{1/(2N+1)-\zeta} n^{r-1} 2^{-ir}\right),$$

where $r>0$ is arbitrary and $\zeta$ is a fixed number in $(0, 1/(2N+1))$. So, for $j=1$ or 3,

$$\begin{aligned}\left(\sum_{i=i_s+1}^{R} T_{3j}^{1/2}\right)^2 &\le C\left[n^{-2s/(2s+1)} + (\log n)\, n^{1/(2N+1)-\zeta+r-1}\left(\sum_{i=i_s+1}^{R} 2^{-ir/2}\right)^2\right]\\ &\le C\left[n^{-2s/(2s+1)} + (\log n)\, n^{1/(2N+1)-\zeta+r-1-r/(2s+1)}\right].\end{aligned}$$

Now,

$$n^{1/(2N+1)-\zeta+r-1-r/(2s+1)} < n^{-2s/(2s+1)}$$

whenever

$$1/(2N+1) < 1/(2s+1) + r/(2s+1) + \zeta - r.$$

Since $s<N$ and $r>0$ is arbitrary, this can be accomplished by taking $r=\zeta>0$. Therefore,

$$\begin{aligned}(\log n)\, n^{1/(2N+1)-\zeta+r-1-r/(2s+1)} &= (\log n)\, n^{1/(2N+1)-1-\zeta/(2s+1)}\\ &\le (\log n)\, n^{-2s/(2s+1)} n^{-\zeta/(2s+1)}\\ &\le Cn^{-2s/(2s+1)},\end{aligned}$$

and

$$\left(\sum_{i=i_s+1}^{R} T_{3j}^{1/2}\right)^2 \le C\left(n^{-2s/(2s+1)}\right).$$

The bounds for $T_{32}$ and $T_{34}$ are unchanged from (34) and (35). Then, the bound for $T_3$ at (41) becomes

$$T_3 \le C\left[n^{-\gamma} + \left(\log_2 n^{1-\zeta}\right)^2 n^{-\delta} + n^{-2s/(2s+1)}\right],$$

where $\gamma$ and $\delta$ are as in the proof of theorem 2.

For the piece $T_4$, we have as in the proof of theorem 2 that

$$\|D_i f\|_2^2 \le C\left(2^{-2si} + n^{1/(2N+1)-\zeta}2^{-i}\right).$$

Repeating the argument at (42),

$$\begin{aligned} T_4 &\le C\left(2^{-2Rs} + n^{1/(2N+1)-\zeta}2^{-R}\right) \\ &\le C\left(n^{-2s(1-\zeta)} + n^{1/(2N+1)-\zeta}n^{-(1-\zeta)}\right) \\ &= C\left(n^{-2s(1-\zeta)} + n^{-2N/(2N+1)}\right) \\ &\le C\left(n^{-2s(1-\zeta)} + n^{-2s/(2s+1)}\right), \end{aligned}$$

for $1/2 < s < N$. Since the other pieces are not affected by the change in the number of discontinuities of the irregular part of $f$, we have

$$\begin{aligned} E\|f-\hat{f}\|_2^2 \le C\Big[&n^{-2s/(2s+1)} + \left(\log_2 n^{1/(2s+1)}\right)^2 n^{-\delta} + n^{-\gamma} \\ &+ \left(\log_2 n^{1-\zeta}\right)^2 n^{-\delta} + n^{-2s(1-\zeta)}\Big]. \end{aligned}$$

This bound will be of order smaller than $n^{-2s/(2s+1)}$ whenever $\zeta \in (0, 1/2)$ and $\gamma \ge 2N/(2N+1)$. This condition on $\gamma$ is satisfied if

$$c = A(0.08)^{-1}\left(C_2\|Q\|_2 + \|Q\|_1\sqrt{\frac{2N}{C_1(2N+1)}}\right)^2.$$

Although $\zeta$ may be as large as 1/2, it does not make sense unless the order of the number of discontinuities is a positive power of $n$. Therefore, the restriction on $\zeta$ is that it lie in the interval $(0, 1/(2N+1)$.

### *4.4 Proof of Thoerem 1*

Recall that the wavelet estimator is

$$\begin{aligned} \hat{f}(x) &= \sum_j \hat{\alpha}_j\phi_j(x) + \sum_{i=0}^{R}\sum_k\sum_{j\in B(k)} \hat{\beta}_{ij}\psi_{ij}(x)I(\hat{B}_{ik} > cn^{-1}) \\ &= \hat{K}_0(x) + \sum_{i=0}^{R}\sum_k \hat{D}_{ik}(x)I(x\in J_{ik})I(\hat{B}_{ik} > cn^{-1}). \end{aligned}$$

The proof in the wavelet case is very similar to that of the convolution case. As before, write

$$\begin{aligned} E\|\hat{f}-f\|_2^2 \le{}& 4E\left\|\hat{K}_0 - K_0\right\|_2^2 + 4E\left\|\sum_{i=0}^{i_s}\left[\sum_k \hat{D}_{ik}I(J_{ik})I(\hat{B}_{ik} > cn^{-1}) - D_i f\right]\right\|_2^2 \\ &+ 4E\left\|\sum_{i=i_s+1}^{R}\left[\sum_k \hat{D}_{ik}I(J_{ik})I(\hat{B}_{ik} > cn^{-1}) - D_i f\right]\right\|_2^2 + 4\left\|\sum_{i=R+1}^{\infty} D_i f\right\|_2^2 \\ ={}& W_1 + W_2 + W_3 + W_4 \end{aligned}$$

Lemma 1 implies the bound on $W_1$ is the same as in the convolution proof,

$$W_1 \leq C/n.$$

To bound $W_2$, the following integral for a fixed $i$ is of use. Using the orthogonality of the wavelet functions,

$$\begin{aligned}
&E\int\left(\sum_k \hat{D}_{ik}(x)I(x\in J_{ik})I(\hat{B}_{ik}>cn^{-1})-D_if(x)\right)^2 dx\\
&=E\int\left(\sum_k\sum_{j\in B(k)}\left(\hat{\beta}_{ij}\psi_{ij}(x)I(\hat{B}_{ik}>cn^{-1})-\beta_{ij}\psi_{ij}(x)\right)\right)^2 dx\\
&=E\int\sum_k\left[\sum_{j\in B(k)}\left(\hat{\beta}_{ij}\psi_{ij}(x)I(\hat{B}_{ik}>cn^{-1})-\beta_{ij}\psi_{ij}(x)\right)\right]^2 dx\\
&=\sum_k E\int_{J_{ik}}\left[\sum_{j\in B(k)}\left(\hat{\beta}_{ij}\psi_{ij}(x)I(\hat{B}_{ik}>cn^{-1})-\beta_{ij}\psi_{ij}(x)\right)\right]^2 dx\\
&=\sum_k E\int_{J_{ik}}\left(\hat{D}_{ik}(x)I(\hat{B}_{ik}>cn^{-1})-D_{ik}f(x)\right)^2 dx.
\end{aligned}$$

Then, in a manner similar to section 4.2,

$$\begin{aligned}
&E\int\left[\sum_k \hat{D}_{ik}(x)I(\hat{B}_{ik}>cn^{-1})-D_if(x)\right]^2 dx\\
&\leq E\int\left(\hat{D}_i(x)-D_if(x)\right)^2 dx\\
&\quad+E\sum_k\int_{J_{ik}}(D_{ik}f(x))^2\,dx\; I(B_{ik}\leq 2cn^{-1})\\
&\quad+E\sum_k\int_{J_{ik}}(D_{ik}f(x))^2\,dx\; I(\hat{B}_{ik}\leq cn^{-1})I(B_{ik}>2cn^{-1})\\
&=W_{21}+W_{22}+W_{23}.
\end{aligned}$$

$W_{21}$ and $W_{22}$ have the same bounds as in the proof of theorem 4,

$$W_{21}, W_{22} \leq C2^i/n.$$

$W_{23}$ will require lemma 3 and Talagrand's theorem. Lemma 3 is unchanged if $J_{ik}$ is substituted for $I_{ik}$, and $D_if$ and $\hat{D}_i$ are replaced by $D_{ik}f$ and $\hat{D}_{ik}$, respectively. For Talagrand's theorem, make the same subsitutions as above, and change $G$ to $G'$, where

$$G'=\left\{\int_{J_{ik}} g(x)D_{ik}(x,\cdot)I(j\in B(k))dx : \|g\|_2\leq 1\right\},$$

and

$$D_{ik}(x,y) = \sum_{j\in B(k)} \psi_{ij}(x)\psi_{ij}(y).$$

Then the same values of $M$, $v$, and $H$ are obtained as in section 4.2. Following the arguments in for $T_{23}$,

$$W_{23} \leq Cn^{-\delta},$$

and

$$W_2 \leq C\left[n^{-2s/(2s+1)} + (\log_2 n^{1/(2s+1)})^2 n^{-\delta}\right],$$

where the value of $\delta$ is the same as in (46).

The bound for $W_3$ is found in a similar manner to Hall, et al. and section 4.2 of this paper. Write

$$\begin{aligned}
&E\int\left[\left(\sum_k \hat{D}_{ik}(x)I(x\in J_{ik})I(\hat{B}_{ik} > cn^{-1})\right) - D_i f(x)\right]^2 dx \\
&\leq E\sum_k \int_{J_{ik}} \left(\hat{D}_{ik}(x) - D_{ik}f(x)\right)^2 dx\ I(\hat{B}_{ik} > cn^{-1})I(B_{ik} > c/(2n)) \\
&\quad + E\sum_k \int_{J_{ik}} \left(\hat{D}_{ik}(x) - D_{ik}f(x)\right)^2 dx\ I(\hat{B}_{ik} > cn^{-1})I(B_{ik} \leq c/(2n)) \\
&\quad + E\sum_k \int_{J_{ik}} (D_{ik}f(x))^2\, dx\ I(\hat{B}_{ik} \leq cn^{-1})I(B_{ik} \leq 2cn^{-1}) \\
&\quad + E\sum_k \int_{J_{ik}} (D_{ik}f(x))^2\, dx\ I(\hat{B}_{ik} \leq cn^{-1})I(B_{ik} > 2cn^{-1}) \\
&= W_{31} + W_{32} + W_{33} + W_{34}.
\end{aligned}$$

$W_{32}$ and $W_{34}$ are bounded with Talagrand's theorem just as $T_{32}$ and $T_{34}$ are at (34) and (35) in section 4.2. As there, it is required that $2^R \leq n^{1-\varepsilon}$. Additionally, the same values of $\gamma$ and $\delta$ are obtained.

$$W_{32} \leq 2^i/\log n\left[n^{-\gamma-1}\log n + \left(n^{-1} + n^{-1}\sqrt{2^i\log n/n} + 2^i n^{-2}\right)\ \exp\left(-\frac{nd}{2^i}\right)\right],$$

$$W_{34} \leq Cn^{-\delta}.$$

For $W_{31}$ and $W_{33}$, the argument from Hall, et al. Hall et al. (1998) is unchanged and results in a bound of

$$W_{31}, W_{33} \leq Cn^{-2s/(2s+1)},$$

provided $s_1 > s$. Therefore,

$$W_3 \leq C\left[n^{-2s/(2s+1)} + n^{-\gamma} + \left(\log_2 n^{1-\varepsilon}\right)^2 n^{-\delta}\right].$$

The final piece $W_4$ is easily bounded:

$$W_4 = C \left\| \sum_{i=R+1}^{\infty} D_i f \right\|_2^2 = C \sum_{i=R+1}^{\infty} \sum_j \beta_{ij}^2. \tag{47}$$

Since $f = f_1 + f_2$ where $f_1 \in B_{2,\infty}^s$ and $f_2 \in B_{(s+1/2)^{-1},\infty}^{s_1} \subset B_{2\infty}^{s_1 - s}$ by (2),

$$\begin{aligned} \beta_{ij} &= \int f(x)\psi_{ij}(x)dx \\ &= \int \left(f_1(x) + f_2(x)\right)\psi_{ij}(x)dx \\ &= \beta_{1ij} + \beta_{2ij}, \end{aligned}$$

and (47) becomes

$$W_4 \le C \left[ \sum_{i=R+1}^{\infty} \sum_j \left(\beta_{1ij}^2 + \beta_{2ij}^2\right) \right].$$

From (4),

$$\sum_j \beta_{1ij}^2 \le C 2^{-2is},$$

and

$$\sum_j \beta_{2ij}^2 \le C 2^{-2i(s_1 - s)}.$$

Therefore,

$$\begin{aligned} W_4 &\le C \left( \sum_{i=R+1}^{\infty} 2^{-2si} + 2^{-2i(s_1 - s)} \right) \\ &\le C \left(2^{-2Rs} + 2^{-2R(s_1 - s)}\right) \\ &\le C \left(n^{-2s(1-\varepsilon)} + n^{-2(s_1 - s)(1-\varepsilon)}\right) \\ &\le C n^{-2s/(2s+1)}, \end{aligned}$$

provided that

$$s_1 - s \ge \frac{s}{(2s+1)(1-\varepsilon)} \text{ and } \varepsilon \le 1/2.$$

So, by choosing $c$ such that

$$\gamma = \frac{C_1 \left(\sqrt{0.08c} - C_2 \sqrt{\|f\|_\infty \|Q\|_2^2}\right)^2}{\|f\|_\infty \|Q\|_1^2} \ge \frac{2N}{2N+1}.$$

or,

$$c = A(0.08)^{-1} \left( C_2 \|Q\|_2 + \|Q\|_1 \sqrt{\frac{2N}{C_1(2N+1)}} \right)^2, \tag{48}$$

the desired bound is attained:

$$E\|\hat{f}-f\|_2^2 \le Cn^{-2s/(2s+1)}.$$

## *4.5 Proof of Thoerem 4*

To simplify the proof, assume that $f$ is in $\Lambda^s(M)$ rather than in the local Hölder clasees $\Lambda^s(M, t_0, \delta)$ for points $t_0$ in the support of $f$. Write $\hat{f}(t_0) - f(t_0)$ as

$$\begin{aligned}
\hat{f}(t_0) - f(t_0) &= \sum_j (\hat{\alpha}_j - \alpha_j)\,\phi_j(t_0) \\
&\quad + \sum_{i=0}^{i_s} \sum_k \sum_{j \in B(k)} \left(\hat{\beta}_{ij}\psi_{ij}(t_0) I(\hat{B}_{ik} > cn^{-1}) - \beta_{ij}\psi_{ij}(t_0)\right) \\
&\quad + \sum_{i=i_s+1}^{R} \sum_k \sum_{j \in B(k)} \left(\hat{\beta}_{ij}\psi_{ij}(t_0) I(\hat{B}_{ik} > cn^{-1}) - \beta_{ij}\psi_{ij}(t_0)\right) \\
&\quad + \sum_{i=R+1}^{\infty} \sum_j \beta_{ij}\psi_{ij}(t_0) \\
&= L_1 + L_2 + L_3 + L_4
\end{aligned}$$

where $i_s$ is as before. Then

$$E\left(\hat{f}(t_0) - f(t_0)\right)^2 \le C\left(EL_1^2 + EL_2^2 + EL_3^2 + EL_4^2\right). \tag{49}$$

In each of these sums, the total number of indices $j$ that intersect the support of $\psi_{ij}$ for any resolution level $i = 0, 1, \ldots, R$ or $\phi_j$ is no more than $2q_0$. This fact will be used several times in the following proof.

*Bound on the linear part $L_1$*

Recalling that $\int \phi^2 = 1$ and that $\phi$ is bounded,

$$\begin{aligned}
EL_1^2 &\le CE\sum_j \left[(\hat{\alpha}_j - \alpha_j)\,\phi_j(t_0)\right]^2 \\
&\le C\|\phi\|_\infty^2 E\sum_j (\hat{\alpha}_j - \alpha_j)^2 \\
&= CE\sum_j \int (\hat{\alpha}_j - \alpha_j)^2 \phi_j^2(x).
\end{aligned}$$

Using the orthogonality of the $\phi_j$,

$$\begin{aligned}
EL_1^2 &\le CE\int \left(\sum_j \hat{\alpha}_j\phi_j(x) - \alpha_j\phi_j(x)\right)^2 \\
&= CE\int \left\{\hat{K}_0(x) - K_0 f(x)\right\}^2 dx.
\end{aligned}$$

By applying lemma 6,

$$EL_1^2 \leq Cn^{-1}. \tag{50}$$

*Bound on the nonlinear part $L_2$*

To bound $L_2$, break it into the following sums:

$$\begin{aligned} EL_2^2 = E &\left[ \sum_{i=0}^{i_s} \sum_k \sum_{j \in B(k)} \left( \hat{\beta}_{ij} - \beta_{ij} \right) \psi_{ij}(t_0) I(\hat{B}_{ik} > cn^{-1}) \right. \\ &\left. + \sum_{i=0}^{i_s} \sum_j \beta_{ij} \psi_{ij}(t_0) I(\hat{B}_{ik} \leq cn^{-1}) \right]^2 \\ &= E(L_{21} + L_{22})^2 \\ &\leq C\left(EL_{21}^2 + EL_{22}^2\right) \end{aligned} \tag{51}$$

To bound $L_{21}$, first apply lemma 1:

$$\begin{aligned} EL_{21}^2 &\leq E \left[ \sum_{i=0}^{i_s} \sum_k \sum_{j \in B(k)} \left( \hat{\beta}_{ij} - \beta_{ij} \right) \psi_{ij}(t_0) I(\hat{B}_{ik} > cn^{-1}) \right]^2 \\ &\leq \left[ \sum_{i=0}^{i_s} \left( E \left[ \sum_k \sum_{j \in B(k)} \left( \hat{\beta}_{ij} - \beta_{ij} \right) \psi_{ij}(t_0) I(\hat{B}_{ik} > cn^{-1}) \right]^2 \right)^{1/2} \right]^2 \\ &\leq \|\psi\|_\infty^2 \left( \sum_{i=0}^{i_s} 2^{i/2} \left[ E \sum_j \left( \hat{\beta}_{ij} - \beta_{ij} \right)^2 \right]^{1/2} \right)^2. \end{aligned} \tag{52}$$

Now, $E\sum_j(\hat{\beta}_{ij} - \beta_{ij})^2$ is of order $n^{-1}$:

$$\begin{aligned} E\sum_j (\hat{\beta}_{ij} - \beta_{ij})^2 &= \sum_j \text{var } \hat{\beta}_{ij} \\ &= \sum_j \text{var } \left( \frac{1}{n} \sum_{m=1}^n \psi_{ij}(X_m) \right) \\ &\leq \sum_j n^{-1} \text{var } \psi_{ij}(X_1) \\ &\leq \sum_j n^{-1} \int \psi_{ij}^2(x) f(x) dx \\ &\leq 2q_0 n^{-1} \|f\|_\infty \int \psi_{ij}^2(x) dx \end{aligned} \tag{53}$$

Using this result, (52) becomes

$$\begin{aligned} EL_{21}^2 &\leq C \left[ \sum_{i=0}^{i_s} 2^{i/2} \left( n^{-1} \right)^{1/2} \right]^2 \\ &\leq Cn^{-2s/(2s+1)}. \end{aligned}$$

The bound for $L_{22}$ is found by breaking it in to two pieces.

$$\begin{aligned} L_{22} &= \sum_{i=0}^{i_s}\sum_k \sum_{j\in B(k)} \beta_{ij}\psi_{ij}(t_0) I(\hat{B}_{ik} \le cn^{-1}) I(B_{ik} > 2cn^{-1}) \\ &\quad + \sum_{i=0}^{i_s}\sum_k \sum_{j\in B(k)} \beta_{ij}\psi_{ij}(t_0) I(\hat{B}_{ik} \le cn^{-1}) I(B_{ik} \le 2cn^{-1}) \qquad (54) \\ &= L_{221} + L_{222}. \end{aligned}$$

Then $EL_{22}^2 \le C\,(EL_{221}^2 + EL_{222}^2)$. The piece $L_{221}$ is bounded using Talagrand's theorem. First, note that by lemma 3 and the fact that

$$f \in \Lambda^s(M) \Rightarrow \sum_j \beta_{ij}^2 \le C2^{-2is},$$

we have

$$\begin{aligned} &E\left(\sum_k \sum_{j\in B(k)} \beta_{ij}\psi_{ij}(t_0) I(\hat{B}_{ik} \le cn^{-1}) I(B_{ik} > 2cn^{-1})\right)^2 \\ &\le CE\left(\sum_k \sum_{j\in B(k)} 2^{-i(s+1/2)} 2^{i/2} \|\psi\|_\infty I(\hat{B}_{ik} \le cn^{-1}) I(B_{ik} > 2cn^{-1})\right)^2 \\ &\le C2^{-2is} E\sum_k \sum_{j\in B(k)} I(\hat{B}_{ik} \le cn^{-1}) I(B_{ik} > 2cn^{-1}) \\ &\le C2^{-2is} \sum_k \sum_{j\in B(k)} P\left(\int \left(\hat{D}_{ik}(x) - D_{ik}f(x)\right)^2 dx > \frac{0.16c\log n}{n}\right). \end{aligned}$$

Using Talagrand's theorem in a manner similar to that of section 4.2,

$$P\left(\int \left(\hat{D}_{ik}(x) - D_{ik}f(x)\right)^2 dx > \frac{0.16c\log n}{n}\right) \le Cn^{-\delta},$$

where $\delta$ is as before. Therefore, using this bound on the probability and lemma 1,

$$\begin{aligned} EL_{221}^2 &\le \left(\sum_{i=0}^{i_s}\left[E\left(\sum_k \sum_{j\in B(k)} \beta_{ij}\psi_{ij}(t_0) I(\hat{B}_{ik} \le cn^{-1}) I(B_{ik} > 2cn^{-1})\right)^2\right]^{1/2}\right)^2 \\ &\le C\left(\sum_{i=0}^{i_s} 2^{-is} n^{-\delta/2}\right)^2 \\ &\le Cn^{-\delta}. \end{aligned}$$

To bound $L_{222}$, observe that orthogonality gives

$$E\left(\sum_k \sum_{j\in B(k)} \beta_{ij}\psi_{ij}(t_0) I(\hat{B}_{ik} \le cn^{-1}) I(B_{ik} \le 2cn^{-1})\right)^2 \le C2^i \|\psi\|_\infty^2 \sum_k \sum_{j\in B(k)} \beta_{ij}^2 I(B_{ik} \le 2cn^{-1}). \tag{55}$$

Now, $B_{ik} \le 2cn^{-1}$ implies that

$$\sum_k \sum_{j\in B(k)} \beta_{ij}^2 \le C \log n/n.$$

By virtue of $f$ being in $\Lambda^s(M)$,

$$\sum_k \sum_{j\in B(k)} \beta_{ij}^2 \le C2^{-2i(s+1/2)}.$$

Therefore,

$$\sum_k \sum_{j\in B(k)} \beta_{ij}^2 I(B_{ik} \le 2cn^{-1}) \le C\left(n^{-1}\log n \wedge 2^{-2i(s+1/2)}\right),$$

and so

$$E\left[\sum_k \sum_{j\in B(k)} \beta_{ij}\psi_{ij}(t_0) I(\hat{B}_{ik} \le cn^{-1}) I(B_{ik} \le 2cn^{-1})\right]^2 \le C2^i\left(n^{-1}\log n \wedge 2^{-2i(s+1/2)}\right)$$

Therefore, the bound on $L_{222}$ is

$$EL_{222}^2 \le \left(\sum_{i=0}^{i_s}\left[E\left(\sum_k \sum_{j\in B(k)} \beta_{ij}\psi_{ij}(t_0) I(\hat{B}_{ik} \le cn^{-1}) I(B_{ik} \le 2cn^{-1})\right)^2\right]^{1/2}\right)^2 \le C\left[\sum_{i=0}^{i_s} 2^{i/2}\left(n^{-1}\log n \wedge 2^{-2i(s+1/2)}\right)^{1/2}\right]^2.$$

Now, $n^{-1}\log n \le 2^{-2i(s+1/2)}$ whenever $2^i \le \left(n(\log n)^{-1}\right)^{1/(2s+1)}$. Therefore, letting $i_*$ be the integer such that $2^{i_*} \le \left(n(\log n)^{-1}\right)^{1/(2s+1)} < 2^{i_*+1}$,

$$\begin{aligned} EL_{222}^2 &\le C\left(\sum_{i=0}^{i_*} 2^{i/2}\sqrt{\log n/n} + \sum_{i=i_*+1}^{i_s} 2^{i/2}2^{-i(s+1/2)}\right)^2 \\ &\le C\log n/n\left(\sum_{i=0}^{i_*} 2^{i/2}\right)^2 + C\left(\sum_{i=i_*+1}^{i_s} 2^{-is}\right)^2 \\ &\le C\frac{\log n}{n}\left(\frac{n}{\log n}\right)^{1/(2s+1)} + C\left(\frac{\log n}{n}\right)^{2s/(2s+1)} \\ &\le C\left(\frac{\log n}{n}\right)^{2s/(2s+1)}. \end{aligned}$$

The bound on $EL_{22}^2$ is therefore

$$C\left(n^{-\delta}+(n^{-1}\log n)^{2s/(2s+1)}\right),$$

and hence

$$EL_2^2 \le C\left(EL_{21}^2+EL_{22}^2\right) \le C\left[n^{-\delta}+\left(n^{-1}\log n\right)^{2s/(2s+1)}\right]. \tag{56}$$

*Bound on the nonlinear part $L_3$*

As with $L_2$, break $L_3$ into the following parts:

$$\begin{aligned} EL_3^2 &= E\left(\sum_{i=i_s+1}^{R}\sum_{k}\sum_{j\in B(k)}\left(\hat{\beta}_{ij}-\beta_{ij}\right)\psi_{ij}(t_0)I(\hat{B}_{ik}>cn^{-1})\right.\\ &\qquad\left.+\sum_{i=i_s+1}^{R}\sum_{k}\sum_{j\in B(k)}\beta_{ij}\psi_{ij}(t_0)I(\hat{B}_{ik}\le cn^{-1})\right)^2\\ &= E(L_{31}+L_{32})^2\\ &\le C\left(EL_{31}^2+EL_{32}^2\right) \end{aligned}$$

Additionally, $L_{31}$ must be divided as well.

$$\begin{aligned} EL_{31}^2 &\le CE\left[\sum_{i=i_s+1}^{R}\sum_{k}\sum_{j\in B(k)}\left(\hat{\beta}_{ij}-\beta_{ij}\right)\psi_{ij}(t_0)I(\hat{B}_{ik}>cn^{-1})I(B_{ik}>cn^{-1}/2)\right]^2\\ &\quad+CE\left[\sum_{i=i_s+1}^{R}\sum_{k}\sum_{j\in B(k)}\left(\hat{\beta}_{ij}-\beta_{ij}\right)\psi_{ij}(t_0)I(\hat{B}_{ik}>cn^{-1})I(B_{ik}\le cn^{-1}/2)\right]^2\\ &= CEL_{311}^2+CEL_{312}^2. \end{aligned}$$

To take care of $L_{311}$, notice that

$$\begin{aligned} &E\left[\sum_{k}\sum_{j\in B(k)}\left(\hat{\beta}_{ij}-\beta_{ij}\right)\psi_{ij}(t_0)I(\hat{B}_{ik}>cn^{-1})I(B_{ik}>cn^{-1}/2)\right]^2\\ &\le C\sum_{k}E\left[\sum_{j\in B(k)}\left(\hat{\beta}_{ij}-\beta_{ij}\right)\psi_{ij}(t_0)\right]^2 I(B_{ik}>cn^{-1}/2)\\ &\le C\sum_{k}2nc^{-1}B_{ik}E\left[\sum_{j\in B(k)}\left(\hat{\beta}_{ij}-\beta_{ij}\right)\psi_{ij}(t_0)\right]^2. \end{aligned}$$

As in (53),

$$\begin{aligned} E\left[\sum_{j\in B(k)}\left(\hat{\beta}_{ij}-\beta_{ij}\right)\psi_{ij}(t_0)\right]^2 &\le 2^i\|\psi\|_\infty^2 E\sum_{j\in B(k)}\left(\hat{\beta}_{ij}-\beta_{ij}\right)^2\\ &\le C2^i/n. \end{aligned}$$

Since

$$B_{ik} = \frac{1}{\log n} \sum_{j \in B(k)} \beta_{ij}^2 \leq C 2^{-2i(s+1/2)},$$

the bound for $EL_{311}^2$ then follows from an application of lemma 1:

$$\begin{aligned} EL_{311}^2 &\leq C \left[ \sum_{i=i_s+1}^{R} \left( \sum_k \frac{2n}{c} \frac{2^i}{n} 2^{-2i(s+1/2)} \right)^{1/2} \right]^2 \\ &\leq C \left( \sum_{i=i_s+1}^{R} 2^{-is} \right)^2 \\ &\leq C n^{-2s/(2s+1)}. \end{aligned}$$

To bound $EL_{312}^2$, Talagrand's theorem will be used as in section 4.2. To begin, note that by lemma 3

$$\begin{aligned} &E \left[ \sum_k \sum_{j \in B(k)} \left( \hat{\beta}_{ij} - \beta_{ij} \right) \psi_{ij}(t_0) I(\hat{B}_{ik} > cn^{-1}) I(B_{ik} \leq cn^{-1}/2) \right]^2 \\ &\leq C 2^i \|\psi\|_\infty^2 E \sum_k \sum_{j \in B(k)} \left( \hat{\beta}_{ij} - \beta_{ij} \right)^2 I(\hat{B}_{ik} > cn^{-1}) I(B_{ik} \leq cn^{-1}/2) \\ &\leq C 2^i E \sum_k \int_{J_{ik}} \left( \hat{D}_{ik}(x) - D_{ik} f(x) \right)^2 dx \\ &\qquad \cdot I \left( \int_{J_{ik}} \left( \hat{D}_{ik}(x) - D_{ik} f(x) \right)^2 dx > \frac{0.08 c \log n}{n} \right) \end{aligned}$$

From (33) and the fact that the number of indices $k$ intersecting the supportof $\psi_{ij}$ is less than or equal to $2q_0/\log n$, this is bounded by

$$C 2^i (\log n)^{-1} \left[ n^{-\gamma-1} \log n + n^{-\gamma-1} + \left( n^{-1} + n^{-1}\sqrt{2^i \log n/n} + 2^i n^{-2} \right) \exp\left( -\frac{nd}{2^i} \right) \right],$$

where $\gamma$ is as in (29) and $d$ is as in (32). Therefore, repeating the argument for the piece $T_{32}$ at (39),

$$EL_{312}^2 \leq C(n^{-\gamma} + n^{-1}).$$

Only $L_{32}$ still needs bounding.

$$
\begin{aligned}
EL_{32}^2 &= E\left(\sum_{i=i_s+1}^{R}\sum_k\sum_{j\in B(k)} \beta_{ij}\psi_{ij}(t_0)I(\hat{B}_{ik}\le cn^{-1})\right)^2 \\
&\le \left(\sum_{i=i_s+1}^{R}\left[E\left(\sum_k\sum_{j\in B(k)} \beta_{ij}\psi_{ij}(t_0)I(\hat{B}_{ik}\le cn^{-1})\right)^2\right]^{1/2}\right)^2 \\
&\le C\left(\sum_{i=i_s+1}^{R}\sum_k\sum_{j\in B(k)} |\beta_{ij}\psi_{ij}(t_0)|\right)^2 \\
&\le C\left(\sum_{i=i_s+1}^{R} 2q_0 2^{i/2}\|\psi\|_\infty 2^{-i(s+1/2)}\right)^2 \\
&\le C\left(\sum_{i=i_s+1}^{R} 2^{-is}\right)^2 \\
&\le Cn^{-2s/(2s+1)}.
\end{aligned}
$$

The bound for $L_3$ is then

$$
EL_3^2 \le C\left(n^{-2s/(2s+1)} + n^{-\gamma}\right). \tag{57}
$$

*Bound on the nonlinear part $L_4$*

$L_4$ is bounded much like $L_{32}$ was. The only difference is the range of the index $i$ and the lack of an indicator function.

$$
\begin{aligned}
EL_4^2 &= E\left(\sum_{i=R+1}^{\infty}\sum_k\sum_{j\in B(k)} \beta_{ij}\psi_{ij}(t_0)\right)^2 \\
&\le C\left(\sum_{i=R+1}^{\infty}\sum_k\sum_{j\in B(k)} |\beta_{ij}\psi_{ij}(t_0)|\right)^2 \\
&\le C\left(\sum_{i=R+1}^{\infty} 2^{-is}\right)^2 \\
&\le Cn^{-(1-\varepsilon)2s}.
\end{aligned} \tag{58}
$$

*Determination of constants $\gamma, \delta$, $\varepsilon$, and $c$*

From the bounds derived at (50), (56), (57), and (58),

$$
\begin{aligned}
E\left(f(t_0) - \hat{f}(t_0)\right)^2 &\le C\left(n^{-\delta} + (\log n/n)^{2s/(2s+1)} + n^{-\gamma} + n^{-2s/(2s+1)} + n^{-2s(1-\varepsilon)}\right) \\
&\le C(\log n/n)^{2s/(2s+1)}
\end{aligned}
$$

if $\gamma$, $\delta$, and $2s(1-\varepsilon)$ are all positive.

Using (29) and (46) and the fact that $\delta > \gamma$, if

$$c > (0.08)^{-1} C_2^2 \|f\|_\infty \|Q\|_2^2 \tag{59}$$

then $\gamma$ and $\delta$ are positive. However, the choice of the threshold $c$ in theorem 1 is larger than the right side of (59), so this requirement is satisfied.

All that is necessary for $2s(1-\varepsilon)$ to be positive is that $\varepsilon < 1$. The conditions of theorem 3 require that $\varepsilon \in (0, 1/2]$, so this restriction is met.

*Block Lengths of Order Larger than* $\log n$

Suppose the block length $l$ in the wavelet estimator (13) is taken to be of order larger than $\log n$, say

$$l = (\log n)^{1+r}$$

for some $r > 0$. Then, assume that $f$ is a function such that equality (to within a constant factor) is attained in the various inequalities in the treatment of $E\left(\hat{f}(t_0) - f(t_0)\right)^2$, i.e.,

$$\begin{aligned} E\left(\hat{f}(t_0) - f(t_0)\right)^2 &= C\left(EL_1^2 + EL_2^2 + EL_3^2 + EL_4^2\right) \\ &= CEL_{222}^2 + \text{various other terms.} \end{aligned}$$

Also, $f$ is a function and $t_0$ a point such that equality (again to within a constsnat factor) rather than inequality is met in

$$\begin{aligned} EL_{222}^2 &= C\left(\sum_{i=0}^{i_s}\left[E\left(\sum_k \sum_{j\in B(k)} \beta_{ij}\psi_{ij}(t_0) I(\hat{B}_{ik} \le cn^{-1}) I(B_{ik} \le 2cn^{-1})\right)^2\right]^{1/2}\right)^2 \\ &= C\left(\sum_{i=0}^{i_s} 2^{i/2}\|\psi\|_\infty \left[\left(\sum_k \sum_{j\in B(k)} \beta_{ij} I(B_{ik} \le 2cn^{-1})\right)^2\right]^{1/2}\right)^2 \\ &= C\left(\sum_{i=0}^{i_s} 2^{i/2} \sum_k \sum_{j\in B(k)} |\beta_{ij}| I(B_{ik} \le 2cn^{-1})\right)^2. \end{aligned}$$

Repeating the argument for the bound for $EL_{222}^2$ earlier in this section with $\log n$ replaced with $(\log n)^{r+1}$,

$$\sum_k \sum_{j\in B(k)} |\beta_{ij}| I(B_{ik} \le 2cn^{-1}) = C\left(n^{-1/2}(\log n)^{(1+r)/2} \wedge 2^{-i(s+1/2)}\right).$$

Letting $i_r$ be the integer such that $2^{i_r} \le \left(n(\log n)^{-1-r}\right)^{1/(2s+1)} < 2^{i_r+1}$,

$$\begin{aligned}
EL_{222}^2 &= C\left(\sum_{i=0}^{i_r} 2^{i/2} n^{-1/2} (\log n)^{(1+r)/2} + \sum_{i=i_r+1}^{i_s} 2^{i/2} 2^{-i(s+1/2)}\right)^2 \\
&= C\left(n^{-1/2} (\log n)^{(r+1)/2} \sum_{i=0}^{i_r} 2^{i/2} + \sum_{i=i_r+1}^{i_s} 2^{-is}\right)^2 \\
&= C\left(n^{-1/2} (\log n)^{(r+1)/2} \left(n(\log n)^{-1-r}\right)^{1/2(2s+1)} + \left(n(\log n)^{-1-r}\right)^{-s/(2s+1)}\right)^2 \\
&= C\left(\frac{\log n}{n}\right)^{2s/(2s+1)} (\log n)^{2sr/(2s+1)}.
\end{aligned}$$

Therefore,

$$EL_{222}^2 > C\,(\log n/n)^{2s/(2s+1)}$$

for any constant $C$ as $n$ gets larger, and so

$$E\left(\hat{f}(t_0) - f(t_0)\right)^2 > C\,(\log n/n)^{2s/(2s+1)}.$$

# References


Brown, L. and Low, M. (1996). A constrained risk inequality with applications to nonparametric functional estimations. *Ann. Statist.* **24** 2384–2398.

Cencov, N. (1962). Evaluation of an unknown distribution density from observations. *Doklady* **3** 1559–1562.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia.

Donoho, D. and Johnstone, I. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika* **81** 425–455.

Donoho, D., Johnstone, I., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24(2)** 508–539.

Doukhan, P. (1988). Formes de toeplitz associees a une analyse multiechelle. *C. R. Acad. Sci. Paris Ser. A* **306** 663–666.

Doukhan, P. and Leon, J. (1990). Deviation quadratique d'estimateurs de densite par projections orthogonales. *C. R. Acad. Sci. Paris Ser. I Math.* **310** 425–430.

Hall, P., Kerkyacharian, G. and Picard, D. (1998). Block threshold rules for curve estimation using kernel and wavelet methods. *Ann. Statist.* **26** 922–942.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1996). *Wavelets, Approximation and Statistical Applications.* Seminar Berlin, Paris.

Kerkyacharian, G. and Picard, D. (1992). Density estimation in Besov spaces. *Statist. Probab. Lett.* **13** 15–24.

MEYER, Y. (1990). *Ondelettes et Opérateurs: I. Ondelettes.* Hermann et Cies, Paris.

PENSKY, M. (1999). Estimation of a smooth density function using Meyer-type wavelets. *Statist. Decisions* **17** 111–123.

TALAGRAND, M. (1989). Isoperimetry and integrability of the sum of independent Banach-space valued random variables. *Ann. Probab.* **17** 1546–1570.

TALAGRAND, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22** 28–76.

VIDAKOVIC, B. (1995). *Statistical Modeling by Wavelets.* John Wiley and Sons, New York.